Research Article

Title: **Evolutionary aspects of a unique internal mitochondrial targeting signal in nuclear-migrated *rps19* of sugar beet (*Beta vulgaris*)**

Authors: Muneyuki Matsunaga, Yoshiya Takahashi, Rika Yui-Kurino, Tetsuo Mikami, and Tomohiko Kubo

Affiliation: Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Kita-ku, Sapporo 060-8589, Japan

Corresponding author: Tomohiko Kubo,
Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Kita-ku, Sapporo 060-8589, Japan
TEL/FAX: +81-11-706-2484
Email: tomohiko@abs.agr.hokudai.ac.jp

Abstract

The endosymbiotic theory postulates that many genes migrated from endosymbionts to the nuclear genomes of their hosts. Some migrated genes lack presequences directing proteins to mitochondria, and their mitochondrial targeting signals appear to be inscribed in the core coding regions as internal targeting signals (ITSs). ITSs may have evolved after sequence transfer to nuclei or ITSs may have pre-existed before sequence transfer. Here, we report the molecular cloning of a sugar beet gene for ribosomal protein S19 (*Rps19*; the first letter is capitalized when the gene is a nuclear gene). We show that sugar beet *Rps19* (*BvRps19*) is an ITS-type gene. Based on amino-acid sequence comparison, dicotyledonous *rps19*s (the first letter is lower-cased when the gene is a mitochondrial gene), such as tobacco *rps19* (*Ntrps19*), resemble an ancestral form of *BvRps19*. We investigated whether differences in amino-acid sequences between *BvRps19* and *Ntrps19* were involved in ITS evolution. Analyses of the intracellular localization of chimaeric GFP-fusion proteins that were transiently expressed in Welsh onion cells showed that *Ntrps19-gfp* was not localized in mitochondria. When several *BvRps19*-type amino acid substitutions, none of which was seen in any other angiosperm *rps19*, were introduced into *Ntrps19-gfp*, the modified *Ntrps19-gfp* became localized in mitochondria, supporting the notion that an ITS in *BvRps19* evolved following sequence transfer to nuclei. Not all of these substitutions were seen in other ITS-type *Rps19*s, suggesting that the ITSs of *Rps19* are diverse.

Keywords: plant mitochondria, gene migration, ribosomal protein, mitochondrial targeting, mitochondrial gene

Abbreviations: *rps10*, gene for ribosomal protein S10 (mitochondrion encoded); *Rps10*, gene for ribosomal protein S10 (nucleus encoded); ORF, open reading frame; mRNA, messenger RNA, cDNA, DNA complementary to RNA; MTS, mitochondrial targeting

signal; ITS, internal targeting signal; *rps19*, gene for ribosomal protein S19 (mitochondrial encoded); *Rps19*, gene for ribosomal protein S19 (nuclear encoded); GFP, green fluorescent protein; RFP, red fluorescent protein; RPS10, ribosomal protein S10; RPS19, ribosomal protein S19; RT-PCR, reverse transcription-polymerase chain reaction; 5' RACE, 5' rapid amplification of cDNA ends; 3' RACE, 3' rapid amplification of cDNA ends; DAPI, 4'-6-diamidino-2-phenylindole; *BvRps19*, *Beta vulgaris Rps19*; UTR, untranslated region; *Ntrps19*, *Nicotiana tabacum rps19*; *Cox2*, gene for cytochrome oxidase subunit II (nuclear encoded); NLS, nuclear localization signal; TOM, translocase of outer mitochondrial membrane.

1. Introduction

Mitochondria were once free-living organisms resembling α-proteobacteria but became cellular organelles after the completion of endosymbiosis with ancestral eukaryotes (Scheffler, 1999). Mitochondrial genomes, therefore, can be considered as relatives of the initial endosymbionts (Timmis et al., 2004); however, genetic information encoded by the current mitochondrial genomes is insufficient to support life as independent organisms because a number of genes have been lost from the mitochondrial genomes (Gray, 1999; Gray et al., 2004). Genes of bacterial origin have been found in nuclear genomes, and the gene products are imported into mitochondria (e.g. Duchene et al., 2005), indicating that some, but not all, of the lost genes have migrated to nuclear genomes during the course of eukaryotic evolution.

Although the gene repertoire of animal mitochondrial genomes is rather constant, which suggests that gene migration events ceased long ago evolutionarily, the gene repertoire of angiosperm mitochondria is varied (reviewed in Kubo and Newton 2008 and Kubo et al. 2011). This variation has been a starting point for exploring gene migration events; for example, loss of ribosomal protein S10 gene (*rps10*) from

mitochondria occurred in at least 26 lineages during angiosperm evolution, and each of these events was possibly complemented by independent gene migration (Adams et al., 2000). Together with many other instances (Kadowaki et al., 1996; Adams and Palmer, 2003), angiosperms are considered to be one of the best models for the study of gene migration.

Angiosperm mitochondrial genes *per se* appear to function improperly when situated in nuclei (Covello and Gray, 1992). One of the reasons for this problem is that their open reading frames (ORFs) have some cytidine residues that are post-transcriptionally converted to uridine within mRNA, a phenomenon called RNA editing (Takenaka et al., 2008). Therefore, when mitochondrial genes were expressed in nuclei, the codons including editing sites would specify amino acids that are inappropriate for proper function. This issue cannot be ignored because developmental abnormalities occurred when unedited mitochondrial-gene products were engineered to target mitochondria in transgenic plants (Hernould et al., 1993). Gene migration was proposed to involve reverse transcription of edited mRNAs (Nugent and Palmer, 1991), which would provide pre-edited ORFs as cDNAs. However, accumulating evidence indicates that the major routes of organelle DNA transfer to nuclear DNA are mediated by DNA (Woischnik and Moraes, 2002; Shahmuradov et al., 2003; Kleine et al., 2009; Ueda and Kadowaki, 2012).

Another reason for improper gene function of nuclear-situated mitochondrial genes is associated with protein import into mitochondria. Because this process involves a mitochondrial targeting signal (MTS), a sequence in the preprotein that is both necessary and sufficient to direct the protein to mitochondria (Neupert, 1997), gene organization of migrated genes has been inspected to find targeting signals. The migrated genes identified to date have been classified into two classes (for review, see Adams and Palmer, 2003). The first class includes genes with an N-terminal presequence that functions as an MTS. The presequence is cleaved upon import into

mitochondria (Neupert, 1997). Since no such presequence is encoded in mitochondrion-resident genes, migrated genes must have acquired presequences after transfer to nuclei. The origin of presequences associated with migrated genes has been proposed to be either duplicated copies of pre-existing mitochondrial targeting genes or unknown DNA sequences (Kadowaki et al., 1996; Adams et al., 2000; Liu et al., 2009). The second class of migrated genes includes genes lacking presequences. In *Arabidopsis thaliana* and rice (*Oryza sativa*), about one-fourth of the nuclear genes coding for mitochondrial ribosomal proteins lack presequences (Bonen and Calixte, 2006). The MTS of presequence-less genes appears to be inscribed in the core coding regions as internal targeting signals (ITSs) (Neupert, 1997).

An ITS-type *Rps10* (hereafter, the first letter of the gene symbol is shown in upper case when the gene is encoded by nuclear genomes, whereas those shown in lower case are genes encoded by mitochondria) was identified in maize (*Zea mays*) (Murcha et al., 2005). On the other hand, *in vitro* translation products of soybean (*Glycine max*) *rps10*, a mitochondrion-resident gene, are not imported into isolated mitochondria (Murcha et al., 2005). After introducing some amino-acid substitutions into soybean *rps10* with reference to maize *Rps10*, the modified soybean RPS10 protein was successfully imported into isolated mitochondria *in vitro* (Murcha et al., 2005), suggesting that the ITS in maize *Rps10* had evolved after the transfer of the non-importable *rps10* sequence to the nuclear genome.

The mitochondrial gene coding for ribosomal protein S19 (*rps19*) is an interesting example of gene migration because at least 39 independent migrations have occurred during angiosperm evolution (Adams et al., 2002) and both presequence-type and ITS-type *Rps19*s have been found (Sanchez et al., 1996; Adams et al., 2002; Liu et al., 2009). The evolution of the ITS in *Rps19* is open to dispute. ITSs are hypothesized to have pre-existed in ancestral mitochondrion-resident *rps19* as latent ITSs. A latent ITS was found from rice by the following experiment: rice *rps19* was tagged with green

fluorescent protein (GFP) gene, and the fusion gene was expressed in nuclei, but the expressed protein localized to mitochondria (Ueda et al., 2008). An ITS-type *Rps19* was found from melon (*Cucumis melo*) and orange (*Citrus sinensis*) (Liu et al., 2009), but no further analysis was conducted. The questions we wanted to address in this study are whether the latent ITS is preserved in other angiosperm *rps19*, and, if so, whether there are any evolutionary relationships between the latent ITS and the ITS in *Rps19*.

We have determined the entire nucleotide sequence of the sugar beet mitochondrial genome, and found that *rps19* had been lost (Kubo, 2000). A close look at the data presented by Adams et al. (2002) indicated that loss of the mitochondrion-resident *rps19* from sugar beet (Caryophyllales) occurred independently from those of melon (Cucurbitales) and orange (Sapindales) (see also Fig. 1), but it was unknown whether loss of *rps19* from sugar beet mitochondria was complemented by a nuclear-migrated copy. We hypothesized that if a nuclear-migrated copy existed in the sugar beet nuclear genome, sugar beet *Rps19* could be a good example for investigating the evolutionary mechanism for gene migration. Here, we show that sugar beet *Rps19* belongs to a class of ITS-type genes and likely evolved a unique ITS after its ancestral *rps19* sequence was transferred to the nuclear genome.

2. Material and methods

2.1 Plant materials

A sugar beet line TK–81mm–O (Kubo et al., 2000) and a tobacco (*Nicotiana tabacum*) cultivar SR-1 were used in this study. Petunias (*Petunia x hybrida*) were purchased from a local market.

2.2 Nucleic acid isolation

Total cellular DNA was isolated from green leaves according to the procedure of Doyle and Doyle (1990). DNA samples were purified by centrifugation in a CsCl-continuous density gradient, if necessary. Total RNA was isolated from green leaves by using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA). Isolated RNA samples were further purified by incubation with RNase-free DNase I (Takara Bio, Ohtsu, Japan) according to the manufacturer's instructions.

2.3 DNA gel blot analysis

Total cellular DNA was digested with restriction endonucleases purchased from Takara Bio and electrophoresed in a 1% agarose gel. Separated DNA fragments were blotted onto a Hybond N+ membrane (GE Healthcare UK, Amersham Place, England). DNA fragments of interest were labeled with alkaline phosphatases using AlkPhos Direct DNA labeling system (GE Healthcare UK). Hybridization was done according to the instruction manual. Signal bands were detected by the exposure to X-ray films.

2.4 Molecular cloning

Preparation of the TK–81mm–O-genomic library used in this study was described in Matsuhira et al. (2007). About 1 x 10^6 recombinant phages were transferred onto Hybond N membrane (GE Healthcare UK) according to the instruction manual. Plaque hybridization was carried out using the same procedure as that used for DNA gel blot analysis. DNA fragments of interest were subcloned into the pBluescript SK+ vector (Stratagene, La Jolla, CA, U.S.A.). Nucleotide sequences were determined using a Li-COR4200L automated DNA sequencer (Li-COR, Lincoln, NE, U.S.A) or an ABI3130 genetic analyzer (Applied Biosystems, Foster City, CA, U.S.A.). Sequence

analysis was conducted using GENETYX (GENETYX CORPORATION, Tokyo, Japan) or Sequencher (Hitachi Software Engineering, Tokyo, Japan). BLAST searches were conducted at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequence alignment was carried out using CLUSTALW (http://clustalw.ddbj.nig.ac.jp/top-j.html) and modified manually. Subcellular localization was predicted using Predotar (http://urgi.versailles.inra.fr/predotar/predotar.html) and Wolf PSORT (http://wolfpsort.org/). The nucleotide sequences obtained in this study were deposited in DDBJ/EMBL/GenBank under accession numbers AB751612, AB751613 and AB752306.

2.5 Reverse transcription-polymerase chain reaction

Reverse transcription (RT) was performed using Superscript III (Invitrogen, Carlsbad, CA, U.S.A.) according to the instruction manual. The subsequent PCR reaction was carried out using BlendTaq (Toyobo Life Science, Osaka, Japan). See also Table S1.

2.6 5'- and 3'-rapid amplification of cDNA ends

Both of the cDNA ends were PCR amplified using a GeneRacer Kit with Super Script III RT (Invitrogen, Carlsbad, CA) according to the instruction manual. See also Table S1.

2.7 Construction of chimaeric green fluorescent protein genes and transient assays

Procedures for PCR amplification to obtain DNA fragments that were to be fused with the 5' to GFP genes are summarized in Table S1 and Fig. S1. Constructed DNA

fragments were inserted into the pTH2 vector (Chiu et al., 1996), in which GFP gene expression is driven by the Cauliflower-Mosaic Virus 35S promoter. Sequence integrity was verified by nucleotide sequencing. The resultant plasmid was introduced into the epidermal cells of Welsh onion sheathes using an IDERA GIE-III particle delivery system (Frontier Science, Ishikari, Japan). Mitochondria were marked by red fluorescent protein (RFP) fluorescence which was expressed by the co-introduced pMt-R plasmid (Arimura and Tsutsumi, 2002), encoding RFP fused with the N-terminal region of F1-ATPase δ-subunit. Nuclei were marked by RFP with nuclear localization signal of SV40 large T-antigen (van der Krol and Chua, 1991) that was expressed from pNc-R plasmid, which was made by modifying pMt-R (see Table S1). Fluorescent signals were observed with a Nikon E600 fluorescence microscope (Nikon, Tokyo, Japan). The images were captured, pseudo-colored and merged using Photoshop (Adobe, San Jose, CA, USA).

3. Results and Discussion

3.1 Molecular cloning of sugar beet Rps19

A nuclear-transferred *Rps19* sequence in sugar beet was detected by DNA gel blot analysis. The probe was a mitochondrion-resident *rps19* sequence that was PCR amplified from petunia total cellular DNA using a pair of primers (see Table S1). The amplicon was labeled and hybridized to a blot containing sugar-beet total cellular DNA. Three bands, a 7.7-kbp *Bam*HI fragment, a 6.2-kbp *Eco*RI fragment and a 2.6-kbp *Hin*dIII fragment, were detected on the blot (Fig. 2A). We cloned the 2.6-kbp *Hin*dIII fragment and found that the fragment contained an ORF showing high homology to plant mitochondrial *rps19* (see Supporting Information-1).

The 5′ and 3′ termini of mRNA containing the *rps19*-like ORF were investigated by

5′ rapid amplification of cDNA ends (5′-RACE) and 3′-RACE, respectively. Assembly of cDNA sequences revealed a continuous sequence of 809 bp, apart from the poly (A) tail (Fig. S2). In this cDNA sequence, an in-frame termination codon was found upstream of the *rps19*-like ORF. Hereafter, we designated this *rps19*-like ORF as *BvRps19* (after *B. vulgaris Rps19*). The lengths of the 5′ untranslated region (UTR) and the 3′ UTR were 271 and 260 bp, respectively.

Whereas the nucleotide sequence of the 3′ UTR matched the 2596-bp genomic sequence, the -271 to -19 sequence of the 5′ UTR was missing, suggesting the interruption of the 5′ UTR by one or more intron(s) in the sugar-beet genome. PCR amplification of sugar-beet total cellular DNA with a pair of primers (see Table S1), corresponding to the 5′ terminus of the cDNA and internal to *BvRps19*, respectively, generated a 2.8-kbp DNA fragment that was larger than that expected from the cDNA sequence. The nucleotide sequence of the 2.8-kbp fragment had an overlapping region with the 2596-bp genomic sequence, and, by assembling the 2.8-kbp sequence and the 2596-bp genomic sequence, a continuous sequence of 3825-bp was revealed (Fig. S2). Comparison of the cDNA- and the genomic sequences revealed that an intron of 2586 bp intervened between the first exon (253 bp) and the second exon (556 bp) (Fig. 2B). The presence of an intron in the 5′ UTR is atypical of angiosperm mitochondrial genes (Kubo and Newton, 2008).

Transcripts of *BvRps19* were detected from flower bud-, leaf- and root total RNA by RT-PCR (Fig. 2C).

3.2 BvRps19 *lacks an N-terminal presequence*

We compared the amino acid sequence of *BvRps19* with those of several plant *rps19* and *Rps19* deduced proteins, the latter of which include presequence-type (soybean, maize, cotton, and *A. thaliana*) and ITS-type (orange and melon) proteins

(Fig. 3). As a result, it became clear that *BvRps19* belongs to the ITS-type *Rps19* class. Moreover, the amino-acid sequences of ITS-type *Rps19* and *rps19* are highly homologous to each other, but rather diverse from those of presequence-type *Rps19*. This tendency is conspicuous in the first 10 amino acids in the sequences as shown in Fig. 2.

3.3 Intracellular localization is different between BvRPS19-GFP and tobacco RPS19-GFP

Although it is impossible to know the nucleotide sequence of an ancestral mitochondrial copy of *BvRps19*, it seems likely that the ancestral copy was very similar to the present dicotyledonous *rps19*, given that migration of *BvRps19* may have occurred after the divergence of the Caryophyllales lineage (see Fig. 2 of Adams et al. 2002; see Fig. 1 of this study). Additionally, amino-acid sequences of dicotyledonous *rps19* are highly conserved (Fig. 3). Therefore, it is reasonable to consider one of the dicotyledonous *rps19* genes as a model of the ancestral *Bvrps19* gene. We chose tobacco *rps19* (DDBJ/EMBL/GenBank ID: BA000042) for further study.

Tobacco *rps19* (*Ntrps19*, after *N. tabacum rps19*) was PCR amplified from leaf cDNA with a pair of primers (see Table S1). Comparing the cDNA sequence with the PCR-amplified genomic sequence, we found five cytidine-to-uridine RNA editing sites, of which three altered the amino acid specificity of the codons (Fig. S3). These edits increased the homology between BvRPS19 and NtRPS19, suggesting the possibility that the migration of *BvRps19* may have involved reverse transcription of mitochondrial mRNA.

We next examined whether *BvRps19* encoded a mitochondrial protein and whether a latent ITS existed in *Ntrps19*. First, the intracellular localization of BvRPS19 and NtRPS19 was predicted *in silico* by two programs. WolfPsort predicted the nuclear

localization of BvRPS19 (score = 7.0) and NtRPS19 (5.0). Predotar's prediction was ‘possibly mitochondrial (0.46)’ and ‘mitochondrial (0.72)’ for BvRPS19 and NtRPS19, respectively.

The intracellular localization of BvRPS19 was tested by expressing a GFP-tagged chimaeric gene. The entire *BvRps19* was fused 5' to the GFP gene to make a *BvRps19-GFP* fusion gene (#1 in Fig. 4; residue 92 [arginine] was replaced with proline to introduce a restriction site). The fusion gene was delivered into epidermal cells of Welsh onion sheath by particle bombardment. Cells expressing the fusion gene were observed by fluorescence microscopy. As a result, green fluorescent signals were localized to small intracellular particles that clearly matched with the red fluorescent signals of MTP-RFP expressed from the co-introduced plasmid pMt-R (Fig. 4A2 and Fig. 4A3).

The entire *Ntrps19* ORF was fused 5' to the GFP gene (#2 in Fig. 4; residue 94 [lysine] was replaced with threonine to introduce a restriction site) and expressed in Welsh onion sheath epidermal cells. The green signals obtained from NtRPS19-GFP co-localized with red signals from nuclear localizing RFP (#2 in Fig. 4). Our GFP-fusion assay failed to find any signs of a latent ITS from *Ntrps19*.

3.4 Modified NtRPS19-GFP localized to mitochondria

Comparing NtRPS19 with BvRPS19, several amino acid substitutions are present, although the two sequences are very similar (Fig. 3). We next examined whether these substitutions are involved in the differential localization between NtRPS19-GFP and BvRPS19-GFP. The fluorescence localization results of a series of GFP-fusion proteins are shown in Fig. 4.

We first focused our analysis on the N-terminal region of BvRPS19 because the translation products of fusion gene #3 (see Fig. 4; the first 37 amino acid residues of

BvRPS19 were fused to GFP) localized to mitochondria. In the first 37 amino acid residues, the differences between BvRPS19 and NtRPS19 are P2S, I6L, and a diverged region (19th-29th residues), whereas the other residues are identical (Fig. 3). We introduced all these changes into NtRPS19-GFP (#4 in Fig. 4); however, the modified NtRPS19-GFP localized in nuclei. As mentioned by Neupert (1997), overall conformation of a preprotein can influence mitochondrial import due to changes in signal accessibility. It was possible that the conformation of NtRPS19 might be associated with the failure in mitochondrial localization. This notion was supported by our data showing that deletion of K90-to-K94 from #4 resulted in mitochondrial localization of the modified fusion protein (#5 in Fig. 4).

Conversely, #6 (Fig. 4), a fusion protein where K90-to-K94 was deleted but otherwise the same as #2, was nuclearly localized. This result indicates that the K90-to-K94 deletion is insufficient for mitochondrial localization of NtRPS19-GFP, perhaps due to the lack of an ITS in NtRPS19.

Of the three differences in the N-terminal 37 amino acid residues (P2S, I6L, and a diverged region [19th-29th residues]), some of these differences may have functional significance and others may be silent polymorphisms. To identify the functionally significant differences, each of the three regions in #5 was reversed compared to that of NtRPS19. Reversion of the 19th-29th residues had no effect on the localization pattern (#7); however, #8 (P2S substitution alone) fusion protein localized in nuclei as well as in mitochondria (Fig. 4). This localization pattern was distinct from that of #1. Also, #9 (I6L substitution alone) fusion protein was localized to nuclei (Fig. 4). We concluded that both P2S and I6L are functionally significant for the N-terminal segment of BvRPS19 to be imported into mitochondria.

Consistent with the result of the localization assay using #4 fusion protein, the introduction of both P2S and I6L to #2 did not alter the localization pattern (#10 in Fig. 4), indicating that an additional change is necessary. We noticed that, whereas the

C-terminal ~14 amino acid residues of *rps19* deduced proteins were fairly conserved, those of presequence-type and ITS-type *Rps19* diverged from *rps19* (Fig. 3). In addition to this finding, whereas D44-C45-S46 is conserved among *rps19* deduced proteins, this amino-acid sequence is not seen in presequence-type and ITS-type *Rps19*s. We tested whether changes in these regions were associated with the mitochondrial localization of BvRPS19. The 81st to 94th residues of #10 were substituted with the corresponding region of BvRPS19 (81st to 92nd residues, see Fig. 3). The resultant fusion gene was termed #11, and its translation products exhibited nuclear, as well as mitochondrial, localizations (Fig. 4). Then, three additional substitutions, D44G, C45S, and S46A, were introduced (#12). Translation products of fusion gene #12 localized to mitochondria (Fig. 4). On the other hand, #13 fusion protein, in which the second and sixth residues were reversed to proline and isoleucine, respectively, was not specifically localized (Fig. 4).

4. Conclusion

BvRps19 is an ITS-type migrated gene. Assuming that *Ntrps19* represents an ancestral form of *BvRps19*, involvement of latent ITS in *BvRps19* migration seems unlikely. Rather, *BvRps19* migration is associated with sequence alterations in four regions (2nd, 6th, 44th-to-46th, and 81st-to-94th residues). These changes are not seen in any dicotyledonous *rps19* deduced proteins identified so far, suggesting that these alterations occurred after sequence transfer to the nuclear genome. It is very likely that these alterations contain some functionally silent polymorphisms. Further analyses for specifying the significant substitutions are infeasible for us because it would require tests of at least additional 32768 fusion proteins (2^{15}, for single and combinations of 13 substitutions).

Compared to the study of maize *Rps10* and soybean *rps10* (Murcha et al. 2005), in

which changes in the N-terminal 20 residues were sufficient for importing soybean *rps10* protein, the alterations associated with *BvRps19* protein importability were scattered throughout the entire protein. One of the reasons for this difference may be associated with the adopted methodology, i.e. *in vitro* assay (*rps10*) and GFP-fusion assay (this study), as pointed out by Logan (2009).

The two regions, the 44th-to-46th and 81st-to-94th residues, are also altered in other nuclear-migrated *Rps19*s. Therefore, it is possible that alterations in the 44th-to-46th and 81st-to-94th residues may be involved in some other *Rps19* migration events. We hypothesize that any alterations in these regions are forbidden in mitochondria because of an unknown constraint that no longer works in nuclei. In the soybean *Cox2* gene product, successful migration involves a reduction in the hydrophobicity of the transmembrane region (Daley et al., 2002). However, the hydrophobicity of BvRPS19, melon RPS19 and orange RPS19 was not necessarily reduced compared to that of *Ntrps19* (Fig. S4). Therefore, functional aspects of these alterations are unclear. Considering the nuclear localization of NtRPS19-GFP, the C-terminus of *rps19* may function as a nuclear localization signal (NLS) due to the lysine and arginine string that resembles a NLS, but it remains unknown why no mitochondrial targeting happened for a chimaeric protein consisting of the N-terminus of BvRPS19 and the C-terminus of NtRPS19, such as #4 (Fig. 4). Recently, a transcription factor ATFS-1, which has both a MTS and a NLS, was reported to be normally imported into mitochondria but localized to nuclei when mitochondrial import efficiency was reduced (Nargund et al. 2012). Therefore, alterations in the C-terminal region might be associated with mitochondrial import efficiency.

The function of P2S and I6L in *Bvrps19* remains obscure. It is unlikely that these substitutions generated a motif that has been defined as a receptor for TOM complex (TOM20) binding, θXXθθ, where θ is any hydrophobic amino acid and X is an aliphatic amino acid with a preference for those with a long side chain (Murcha et al.,

2005). Melon- and orange *Rps19* have P2S and I6L, respectively (Fig. 3), suggesting that these substitutions also involved in other migration events; however, neither of the two *Rps19* has both substitutions simultaneously. Instead, each of the two *rps19* has unique amino-acid substitutions compared to the dicotyledonous *rps19*, such as R3P and S10G (melon), and L17A (orange). This result led us to infer that the ITS of *Rps19* may be diverse, although further analyses will be necessary to describe fully ITS evolution in dicot *Rps19*s.

Supplementary materials related to this article can be found at the journal online.

Acknowledgements

We would like to thank the DNA Sequencing Facility of the Research Faculty of Agriculture, Hokkaido University, for technical assistance. This work was done in part at the Research Center for Molecular Genetics, Hokkaido University, and was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, and the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN).

References

Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J. and Palmer, J.D., 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature 408, 354-357.

Adams, K.L., Qiu, Y.L., Stoutemyer, M. and Palmer, J.D., 2002. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc. Natl. Acad. Sci. U.S.A. 99, 9905-9912.

Adams, K.L. and Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380-395.

Arimura, S. and Tsutsumi, N., 2002 A dynamin-like protein (ADL2b), rather than FtsZ, is involved in Arabidopsis mitochondrial division. Proc. Natl. Acad. Sci. U.S.A. 99, 5727-5731.

Bonen, L. and Calixte, S., 2006. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. Mol. Biol. Evol. 23, 701-712.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J., 1996. Engineering GFP as a vital reporter in plants. Curr. Biol. 6, 325-330.

Covello, P.S. and Gray, M.W., 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome *c* oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. EMBO J. 11, 3815-3820.

Daley, D.O., Clifton, R. and Whelan, J., 2002. Intracellular gene transfer: Reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome c oxidase. Proc. Natl. Acad. Sci. U.S.A. 99, 10510-10515.

Doyle, J.J. and Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13-15.

Duchêne, A.-M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Maréchal-Drouard, L. and Small, I.D., 2005. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. 102, 16484-16489.

Gray, M.W., 1999. Evolution of organellar genomes. Curr. Opin. Genet. Dev. 9, 678-687.

Gray, M.W., Lang, B.F. and Burger, G., 2004. Mitochondria of protists. Ann. Rev. Genet. 38, 477-524.

Hernould, M., Suharsono, S., Litvak, S., Araya, A. and Mouras, A., 1993. Male-sterility induction in transgenic tobacco plants with an unedited *atp9* mitochondrial gene from wheat. Proc. Natl. Acad. Sci. U.S.A. 90, 2370-2374.

Kadowaki, K.I., Kubo, N., Ozawa, K. and Hirai, A., 1996. Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J. 15, 6652-6661.

Kleine, T., Maier, U.G. and Leister, D., 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Ann. Rev. Plant Biol. 60, 115-138.

Kubo, T., Kitazaki, K., Matsunaga, M., Kagami, H. and Mikami, T., 2011. Male sterility-inducing mitochondrial genomes: how do they differ? Crit. Rev. Plant Sci. 30, 378-400.

Kubo, T. and Newton, K.J., 2008. Angiosperm mitochondrial genomes and mutations. Mitochondrion 8, 5-14.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A. and Mikami, T., 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$(GCA). Nucl. Acids Res. 28, 2571-2576.

Liu, S.-L., Zhuang, Y., Zhang, P. and Adams, K.L., 2009. Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. Mol. Biol. Evol. 26, 875-891.

Logan, D.C., 2009. Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. J. Exp. Bot. 60, 712-714.

Murcha, M.W., Rudhe, C., Elhafez, D., Adams, K.L., Daley, D.O. and Whelan, J., 2005. Adaptations required for mitochondrial import following mitochondrial to nucleus gene transfer of ribosomal protein S10. Plant Physiol. 138, 2134-2144.

Nargund, A.M., Pellegrino, M.W., Fiorese, C.J., Baker, B.M. and Haynes, C.M., 2012. Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation. Science 337, 587-590.

Neupert, W., 1997. Protein import into mitochondria. Ann. Rev. Biochem. 66, 863-917.

Nugent, J.M. and Palmer, J.D., 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66, 473-481.

Sanchez, H., Fester, T., Kloska, S., Schroder, W. and Schuster, W., 1996. Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. EMBO J. 15, 2138-2149.

Scheffler, I.E., 1999. Mitochondria, Wiley-Liss, New York.

Shahmuradov, I.A., Akbarova, Y.Y., Solovyev, V.V. and Aliyev, J.A., 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. Plant Mol. Biol. 52, 923-934.

Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A. and Brennicke, A., 2008. The process of RNA editing in plant mitochondria. Mitochondrion 8, 35-46.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W., 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5, 123-135.

Ueda, M., Fujimoto, M., Arimura, S.-i., Tsutsumi, N. and Kadowaki, K.-i., 2008. Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. Mol. Biol. Evol. 25, 1791-1793.

Ueda, M. and Kadowaki, K.-i., 2012. Gene content and gene transfer from mitochondria to the nucleus during evolution, in: Marechal-Drouard, L. (ed), Advances in Botanical Research vol. 63 Mitochondrial Genome Evolution. Academic Press, Oxford, pp. 21-40.

van der Krol, A.R. and Chua, N.H., 1991. The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. Plant Cell 3, 667-675.

Woischnik, M. and Moraes, C.T., 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res. 12, 885-893.

Figure legends

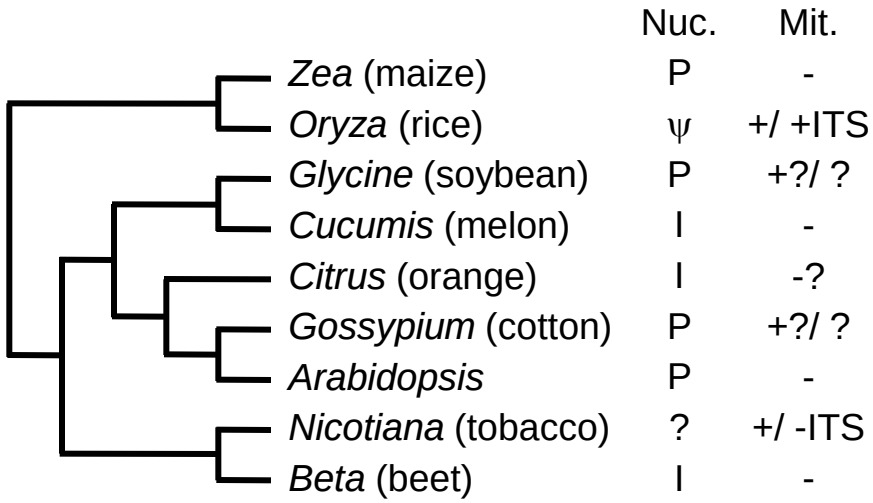
Fig. 1 Phylogenetic relationship and presence/absence of *rps19* and *Rps19* in nine plants. Nuclear (Nuc.) *Rps19* status is abbreviated as P (presence of presequence type), ψ (pseudo gene), I (presence of ITS type) and ? (unknown). Mitochondrial (Mit.) *rps19* status is abbreviated as - (absence), +/ +ITS (presence with latent ITS), -? (probably absence), - (absence), +?/ ? (probably presence but latent ITS unknown), +/ - (presence with no latent ITS). The phylogenetic tree was adopted from the Angiosperm Phylogeny Website (http://www.mobot.org/mobot/research/apweb/welcome.html), but the length of the branches are not correlated with phylogenetic distances. Data from: maize, Adams et al. (2002), DDBJ/GenBank/EMBL accession number EE292591 and AY506529; rice, Ueda et al. (2008) and our unpublished result; soybean, Adams et al. (2002); melon, Adams et al. (2002), Liu et al. (2009), JF412792 and JF412800; orange, Adams et al. (2002) and Liu et al. (2009); cotton, Adams et al. (2002); *Arabidopsis*, Sanchez et al. (1996), and Y08501; tobacco, Adams et al. (2002) and BA000042; sugar beet, BA000009 and this study.

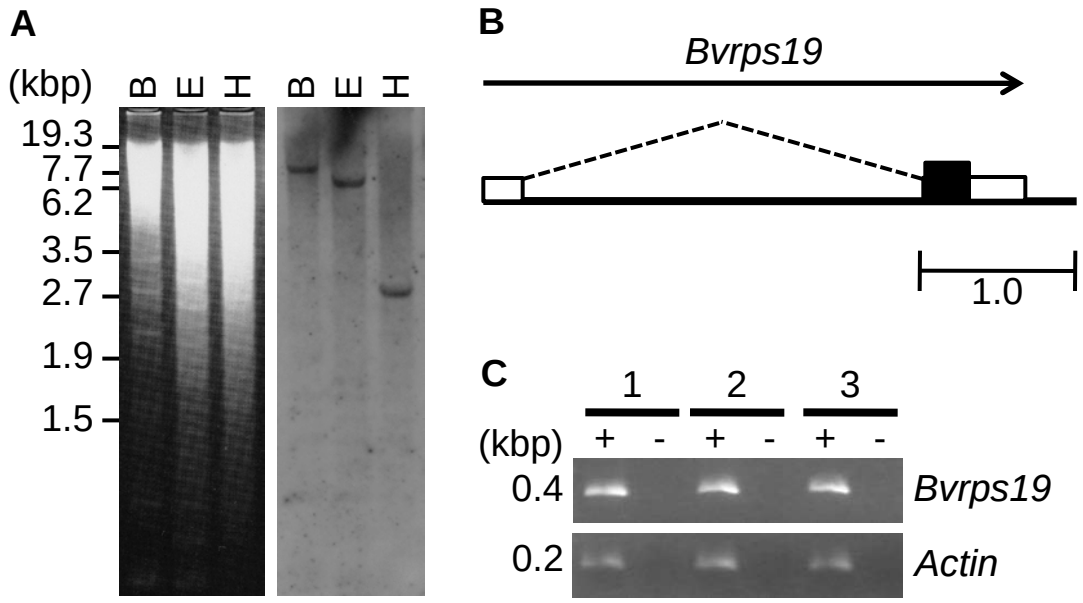
Fig. 2 Organization and expression of sugar beet *rps19* genes. Panel A DNA gel blot analysis of total cellular DNA (5μg) from sugar beet cv. TK–81mm–O probed with petunia *rps19*. Restriction endonucleases used in this experiment are *Bam*HI (B), *Eco*RI (E), and *Hin*dIII (H). Size markers are shown on the left (kb). Panel B Schematic organization of sugar beet *Rps19*. The arrow indicates transcriptional direction. Open- and filled boxes show the UTR and coding region, respectively. An intron interrupts the 5' UTR (depicted by a dotted line). The scale bar is shown in kbp. Panel C RT-PCR analysis of *BvRps19* and the sugar beet *Actin* gene, as a control. Sizes of the PCR products are shown on the left. Total RNAs from flower buds (1), leaves (2), and roots (3) were subjected to reverse transcription with (+) or without (–) reverse transcriptase.

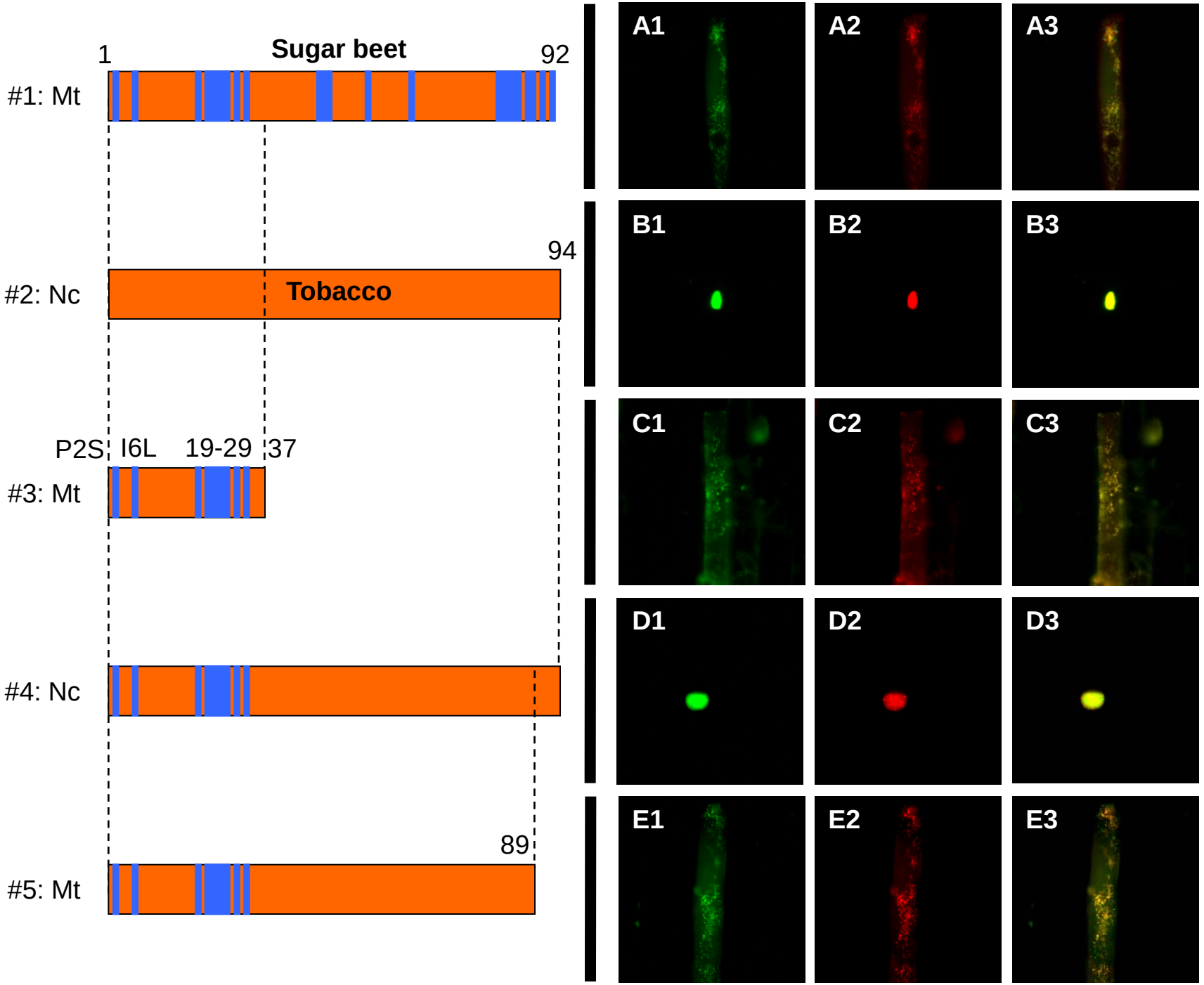
Fig. 3 Alignment of deduced amino acid sequences of RPS19s. Thirteen RPS19 deduced protein sequences were aligned. Residues highlighted in Black indicate amino-acid residues conserved in 7 or more sequences. Positions of amino acid residues that are associated with the ITS evolution of *BvRps19* are highlighted in blue. Nucleus-resident copies and mitochondrion-resident copies are indicated by nc and mt, respectively. Amino acid sequences of mitochondrial RPS19s were deduced from their cDNA sequences. Presequences are not shown, but the lengths of the presequences are shown in parentheses. Sources of sequences are: sugar beet (Bv), this study; tobacco (Nt), DDBJ/GenBank/EMBL accession number BA000042 and this study; grapevine (Vv), RNA editing database (REDIdb) (http://biologia.unical.it/py_script/search.html) accession number EDI_000000871; *Magnolia x soulangeana* (Ms), REDIdb accession number EDI_000000722; rice (Os), REDIdb accession number EDI_000000941; *Cycas revoluta* (Cr), REDIdb accession number EDI_000000253; *Ginkgo biloba* (Gb), REDIdb accession number EDI_000000929; orange (Cs), DDBJ/GenBank/EMBL accession numbers CV718812 and CX071875; melon (Cm), DDBJ/GenBank/EMBL accession number AM727919; soybean (Gm), DDBJ/GenBank/EMBL accession numbers EV279429 and EH260323; maize (Zm), DDBJ/GenBank/EMBL accession number EE292591; cotton (*Gossypium raimondii*), DDBJ/GenBank/EMBL accession number CO082119; *Arabidopsis thaliana* (At), DDBJ/GenBank/EMBL accession number AY049255.

Fig. 4 Summary of localization analyses. A schematic diagram of proteins that were fused with GFP is shown (#1 to #13). Localization of the fusion proteins are abbreviated as: Mt, mitochondria; Nc, nuclei; Mt/Nc, mitochondria and nuclei; nc, no specific localization. The coding region of *Ntrps19* is indicated by orange rectangles. Amino acid substitutions in *BvRps19* are depicted as blue vertical lines. Numbers above the rectangles indicate positions of amino acid residues. The locations of P2S and I6L are

emphatically shown. A1 to O3 are fluorescence images of epidermal cells of Welsh onion sheathes: expressing #1 (A1), and pMt-R (A2); expressing #2 (B1), and pNc-R (B2); expressing #3 (C1), and pMt-R (C2); expressing #4 (D1), and pNc-R (D2); expressing #6 (F1), and pNc-R (F2); expressing #7 (G1), and pMt-R (G2); expressing #8 (H1), and pMt-R (H2); expressing #8 (I1), and pNc-R (I2); expressing #9 (J1), and pNc-R (J2); expressing #10 (K1), and pNc-R (K2); expressing #11 (L1), and pMt-R (L2); expressing #11 (M1), and pNc-R (M2); expressing #12 (N1), and pMt-R (N2); expressing #13 (O1), and pMt-R (O2). Images with a suffix of 3 are merged images of those with suffixes 1 and 2 having the same prefix.


Nuc.
Mit.
Zea (maize)
P
-
Oryza (rice)
ψ
+/ +ITS
Glycine (soybean)
P
+?/ ?
Cucumis (melon)
I
-
Citrus (orange)
I
-?
Gossypium (cotton)
P
+?/ ?
Arabidopsis
P
-
Nicotiana (tobacco)
?
+/ -ITS
Beta (beet)
I
-


A
(kbp)
B E H B E H
19.3
7.7
6.2
3.5
2.7
1.9
1.5
B
Bvrps19
1.0
C
1 2 3
(kbp) + - + - + -
0.4
Bvrps19
0.2
Actin


1
Sugar beet
92
#1: Mt
94
#2: Nc
Tobacco
P2S
I6L
19-29
37
#3: Mt
#4: Nc
89
#5: Mt
A1
A2
A3
B1
B2
B3
C1
C2
C3
D1
D2
D3
E1
E2
E3

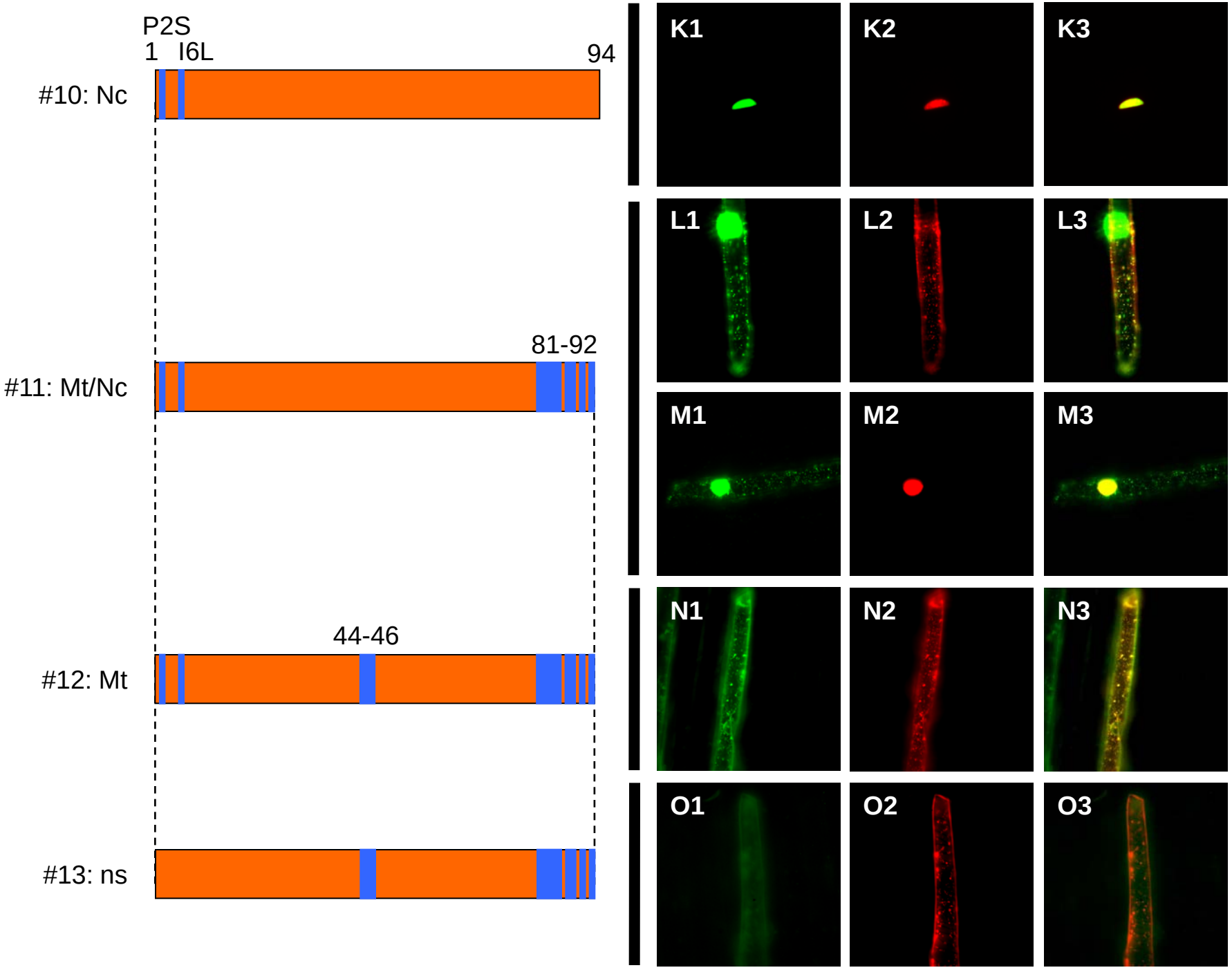
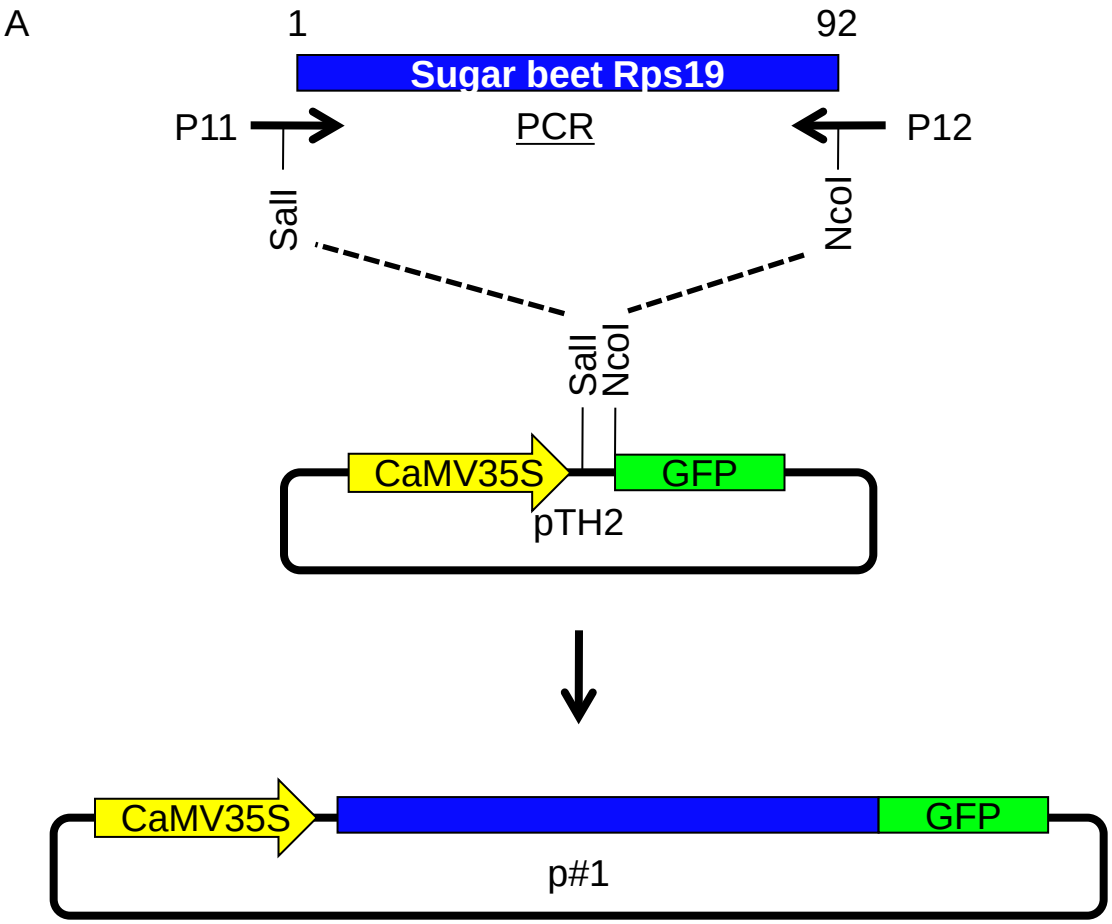
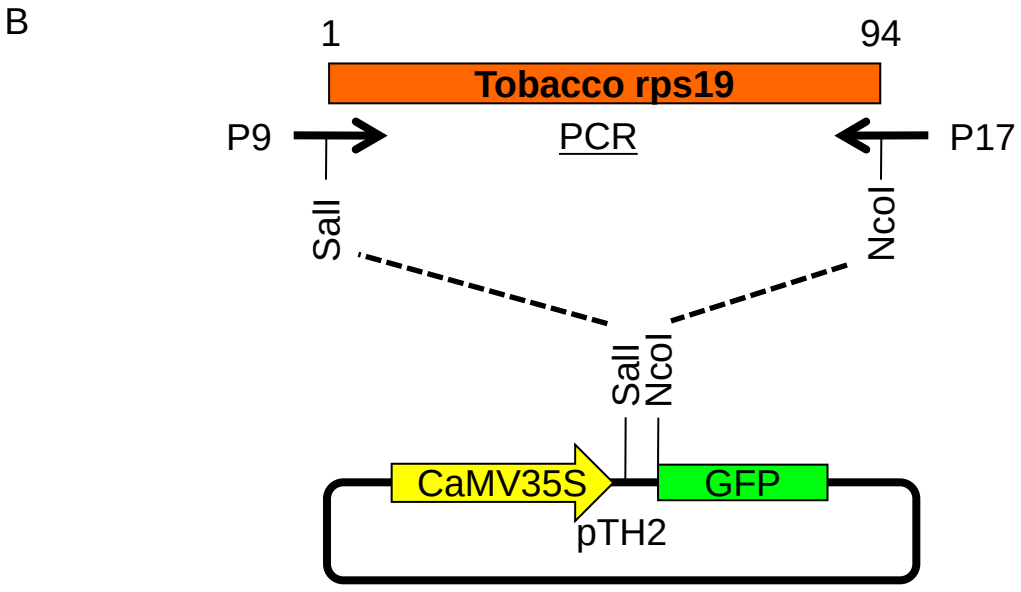
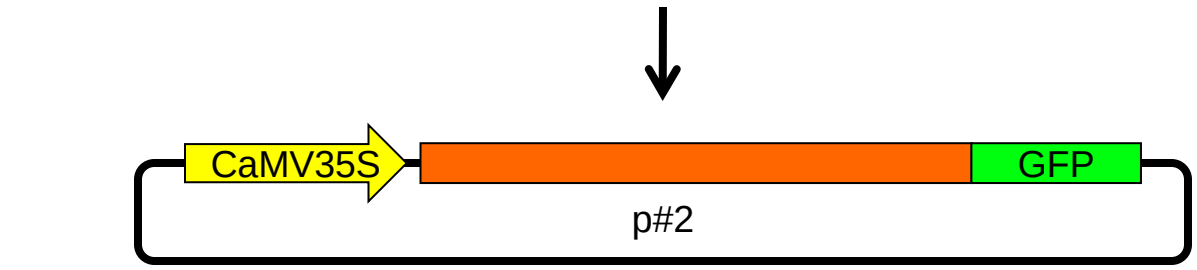

P2S
1
I6L
94
#10: Nc
81-92
#11: Mt/Nc
44-46
#12: Mt
#13: ns
K1
K2
K3
L1
L2
L3
M1
M2
M3
N1
N2
N3
O1
O2
O3

Table S1 Procedures for PCR and primer sequences.

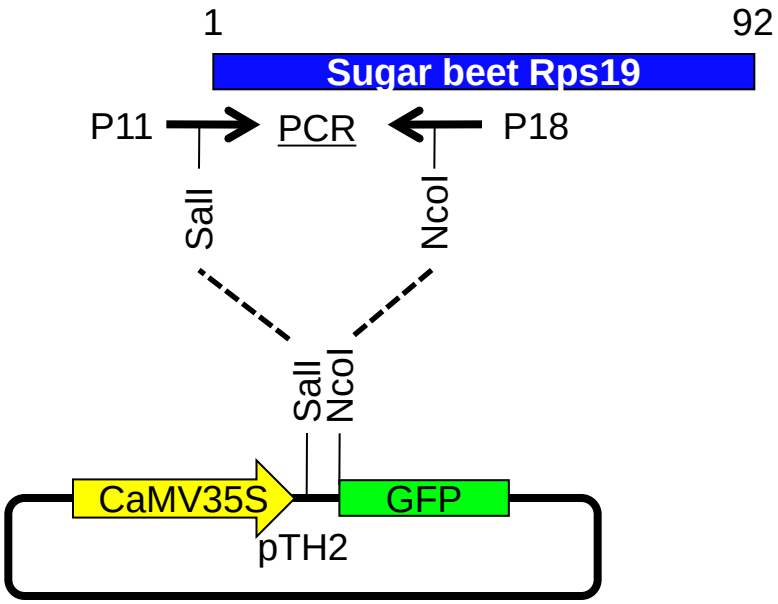
Generation of DNA fragments for hybridization probe	Total cellular DNA of petunia was PCR amplified using P1 (5'-ATGCCACGACGATCCATATG-3') and P2 (5'-GGTCCAATATTTGTTCTCGA-3') primers.
5' RACE	One micro-gram of total cellular RNA from leaves were dephosphorylated and reverse transcribed with GeneRacer Oligo dT Primer (supplied with SUperScript III RT Module) as the manufacturer's instruction. The resultant cDNA was PCR amplified using primers P3 (5'-GAGGCCTTCTTTTCCGTGTG-3') and GeneRacer 5' Primer (supplied by the manufacturer).
3' RACE	The cDNA mentioned above was PCR amplified using primers P4 (5'-AGTCGCACACCGCACCACTC-3') and GeneRacer 3' Primer (supplied by the manufacturer).
Amplification of intronic sequence	Total cellular DNA of sugar beet was amplified using P5 (5'-CAGCGAACGAATGACTTTCCATTG-3') and P6 (5'-GACTTCTCAATTCTCTCTCCTCTCTCCT-3') primers.
RT-PCR	Primers for BvRps19: P5 and P6. Primers for sugar beet actin: P7 (5'-AGACCTTCAATGTGCCTGCT-3') and P8 (5'-ACGACCAGCAAGATCCAAAC-3'). Primers for tobacco rps19: P9 (5'-TCGAAGCGTCGACATTAGTTCATG-3') and P10 (5'-CGTGCCATATGCGCTTAGACTTTA-3').
Tobacco genomic rps19 amplification	Total cellular DNA of tobacco was PCR amplified using P9 and P10.
Construction of fusion gene #1	Plasmid DNA bearing BvRps19 was PCR amplified using P11 (5'-AACGTCGACAATTCCGGAATCATG-3' and P12 (5'-TCACCATGGGACCCCCCTTTTTCT-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #2	Plasmid DNA bearing Ntrps19 cDNA was PCR amplified using P9 and P17 (5'-TCCCATGGTTTTCCCCTTTTTTCTTCC-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #3	Plasmid DNA bearing BvRps19 was PCR amplified using P11 and P18 (5'-CCCATGGAAGACCTCCTTGACCAGA-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #4	Plasmid DNA bearing tobacco rps19 cDNA was PCR amplified using P9 and P19 (5'-CCACTGCAGAACCATGGTTTTCCCC-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and PstI to generate cohesive termini, and then inserted into Sall-PstI digested pBluescript SK+ vector (pNt). Total cellular DNA of sugar beet was amplified with P11 and P20 (5'-CCAACGAATTCTGGCAAGATAGA-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and EcoRI to generate cohesive termini (fragment A). Fragment A was inserted into Sall-EcoRI digested pNt, where EcoRI exists inside of Ntrps19 coding region (pBv-Nt). Note that the differences of N-terminal region between BvNtRPS19 and BvRPS19 are P2S, I6L, and 19th-29th region (see Fig. 2). Sall-NcoI fragment of pBv-Nt was inserted into pTH2.
Construction of fusion gene #5	Plasmid DNA bearing #4 was PCR amplified using P9 and P21 (5'-AACCATGGTTCTTCCCGGTCCAAT-3'). The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #6	Plasmid DNA bearing Ntrps19 cDNA was PCR amplified using P9 and P21. The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #7	Plasmid DNA bearing #7 was PCR amplified using P22 (5'-TCGAAGCGTCGACATTAGTTCATGTCTCGACGATCTTTATGGAAG-3') and P21. The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #8	Plasmid DNA bearing #6 was PCR amplified using P23 (5'-TCGAAGCGTCGACATTAGTTCATGTCTCGACGATCTAT-3') and P21. The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #9	Plasmid DNA bearing #6 was PCR amplified using P24 (5'-TCGAAGCGTCGACATTAGTTCATGCCACGACGATCTTTATGGAAG-3') and P21. The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector.
Construction of fusion gene #10	Plasmid DNA bearing #2 was PCR amplified using P22 and P17. The resultant PCR fragments were digested with restriction endonucleases Sal I and NcoI to generate cohesive termini, and then inserted
Construction of fusion gene #11	Plasmid DNA bearing #10 was PCR amplified using P22 and P25 (5'-ACCATGGGACCCCCCTTTTTCTTTCCAATATTTGTTCTCGAAGGTCTTCG-3'). The resultant PCR fragments were digested with restriction endomucleases Sal I and Nco I to generate cohesive termini, and then inserted into Sall-NcoI digested pTH2 vector (p#11.5). The plasmid DNA was PCR amplified using P26 (5'-CGGAAACGAAGACCTCACAAATTAGATGGTGGAAAGAAAAAGGGG-3') and P27 (5'-CCCCTTTTTCTTTCCACCATCTAATTTGTGAGGTCTTCGTTTCCG-3'). The resultant mixture was digested with Dpn I, which would digest methylated-plasmid DNA template but PCR products would remain intact. Competent *E. coli* cells were transformed with the digest, which contained mutated plasmid wih nick. The nick is sealed in the *E. coli* cells.
Construction of fusion gene #12	Plasmid DNA bearing #11 was PCR amplified using P28 (5'-TGCCGGAATTCGTTGGGTAGCGCTTACGAATTTACAATG-3') and P29 (5'-CATTGTAAATTCGTAAGCGCTACCCAACGAATTCCGGCA-3'). The resultant mixture was digested with Dpn I. Competent E. coli cells were transformed with the digest.
Construction of fusion gene #13	Sall-EcoRI region of p#12 was replaced with that of pNt.
Construction of pNc-R	pMt-R DNA was PCR amplified using P30 (5'-ACAATTACAGTCGACATGGCTCCCAAGAAGAAGAGAAAGGTAATGGTGCGCTC TCC-3') and P31 (5'-GGAGGAGCGCACCATTACCTTTCTCTTCTTCTTGGGAGCCATGTCGACTGTAATTGT-3'). The resultant mixture was digested with Dpn I. Competent E. coli cells were transformed with the digest.

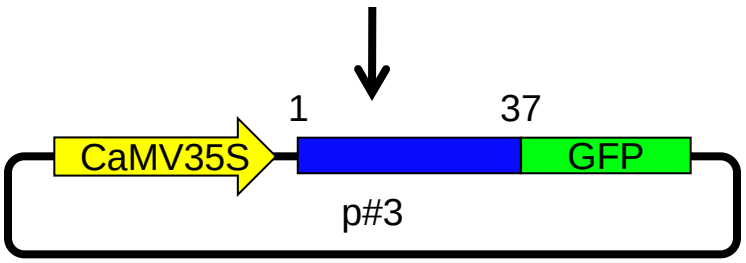

A
1
92
Sugar beet Rps19
P11
PCR
P12
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
CaMV35S
GFP
p#1


B
1
94
Tobacco rps19
P9
PCR
P17
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2

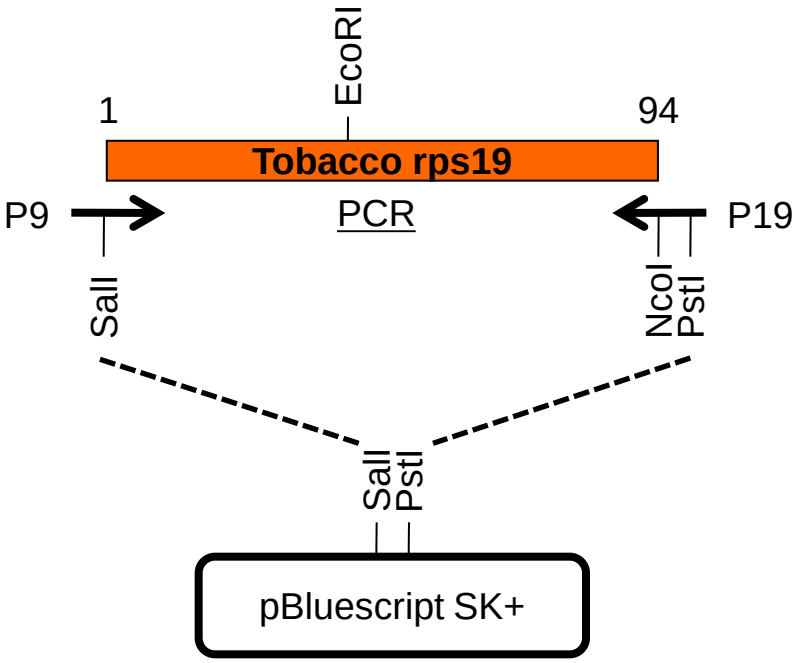

CaMV35S
GFP
p#2

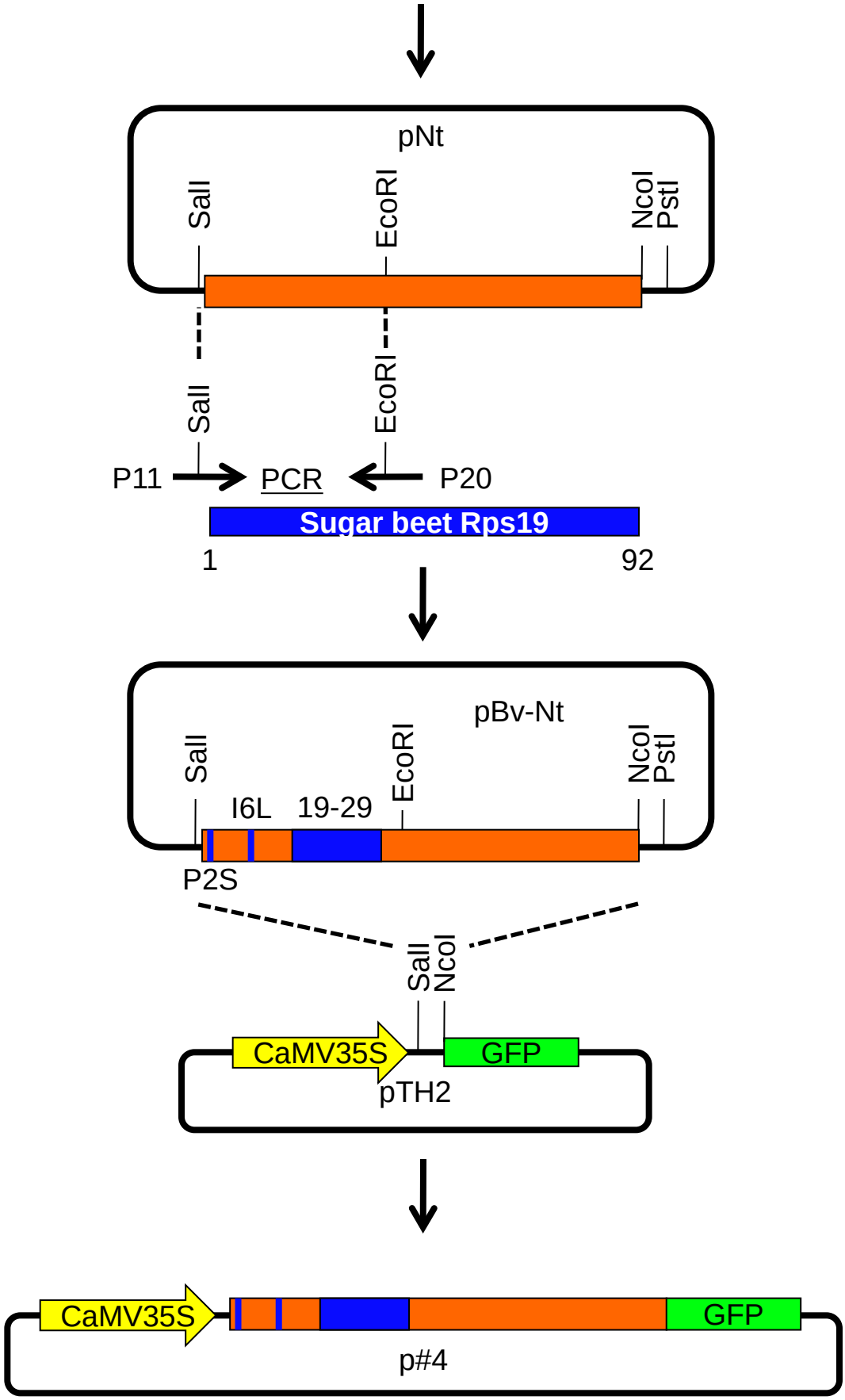
C


1
92
Sugar beet Rps19
P11
PCR
P18
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2

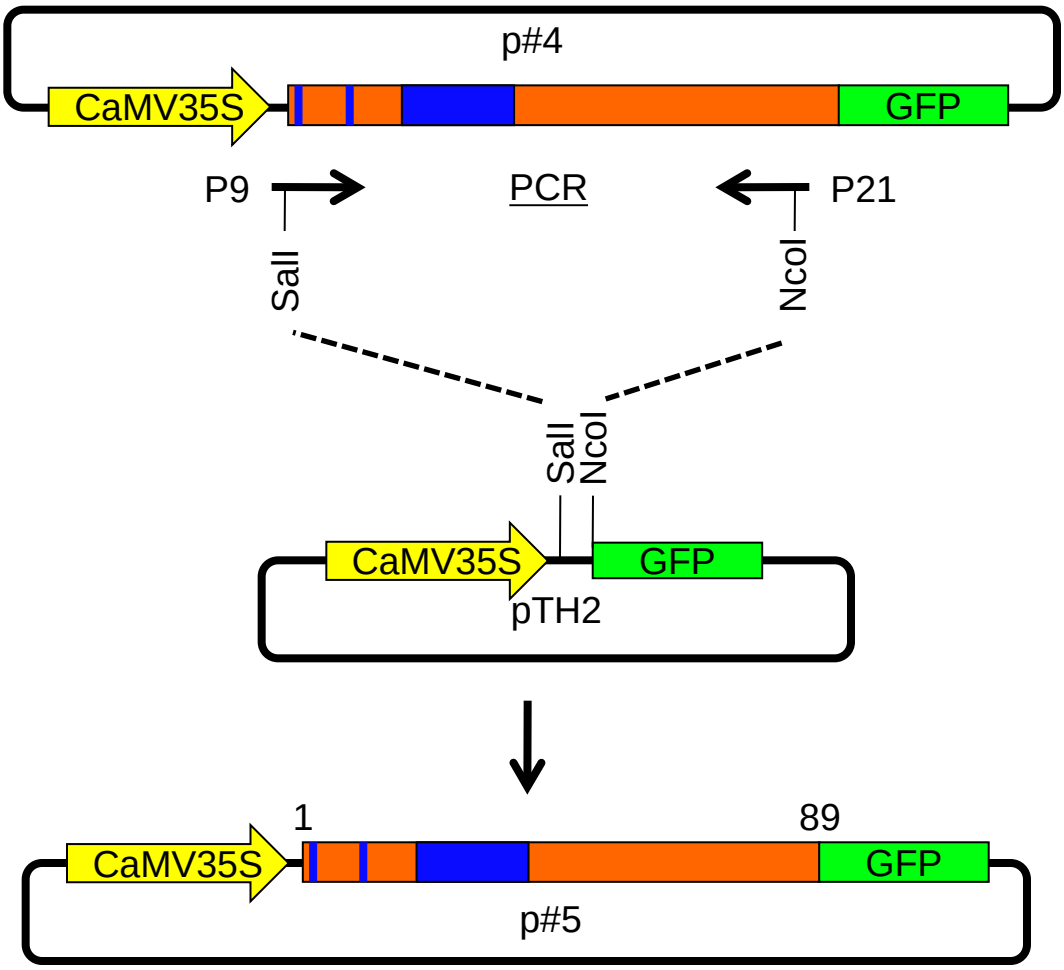

1
37
CaMV35S
GFP
p#3

D

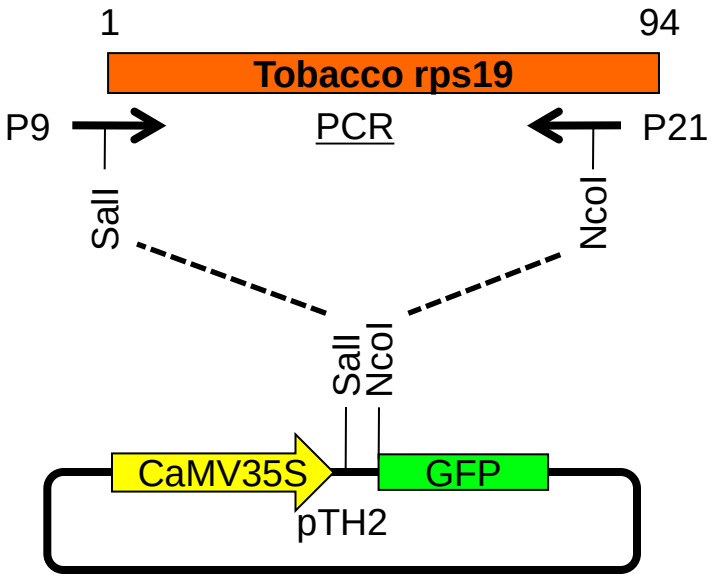

EcoRI
1
94
Tobacco rps19
P9
PCR
P19
SalI
NcoI
PstI
SalI
PstI
pBluescript SK+

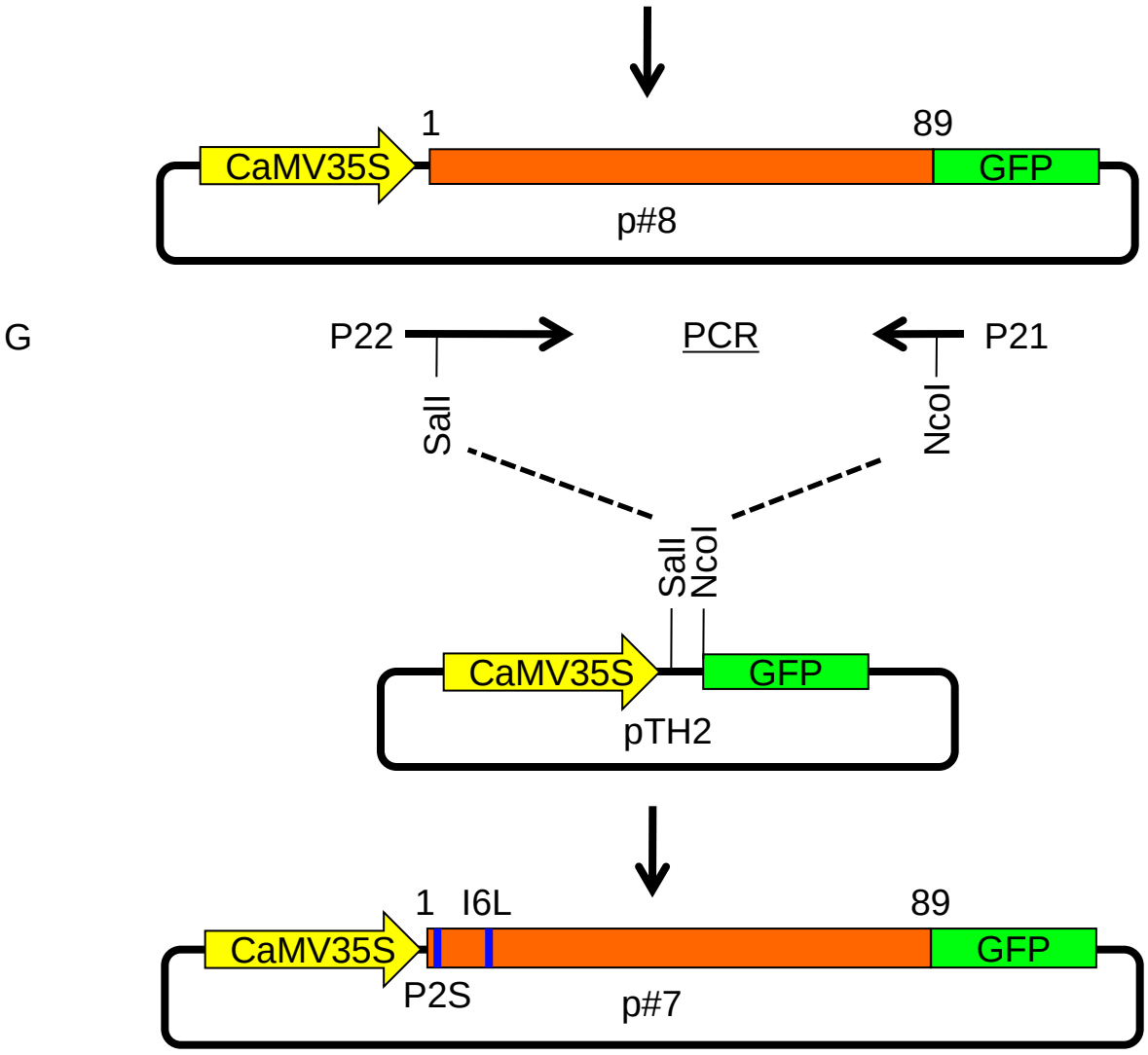

pNt
SalI
EcoRI
NcoI
PstI
SalI
EcoRI
P11
PCR
P20
Sugar beet Rps19
1
92
pBv-Nt
SalI
EcoRI
NcoI
PstI
I6L
19-29
P2S
SalI
NcoI
CaMV35S
GFP
pTH2
CaMV35S
GFP
p#4

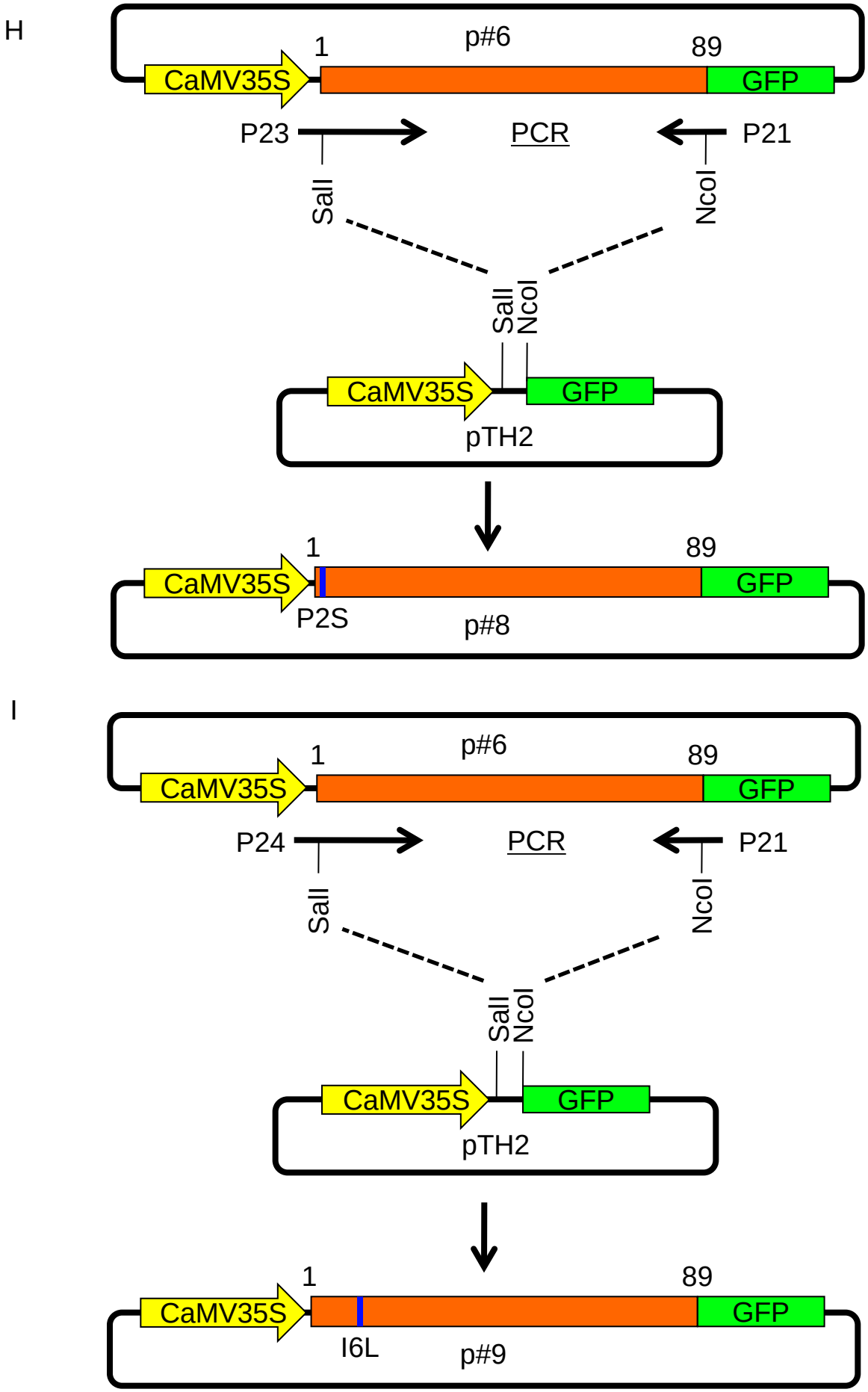
E


p#4
CaMV35S
GFP
P9
PCR
P21
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
89
CaMV35S
GFP
p#5

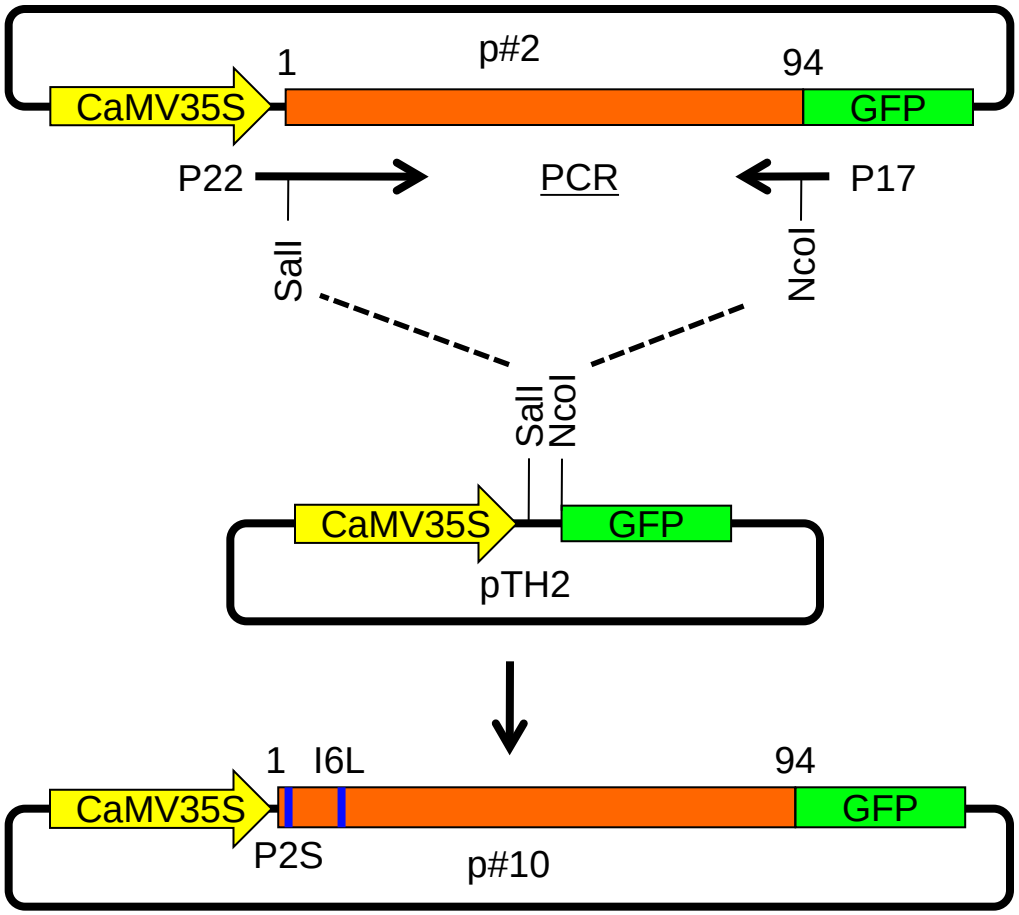
F


1
94
Tobacco rps19
P9
PCR
P21
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2

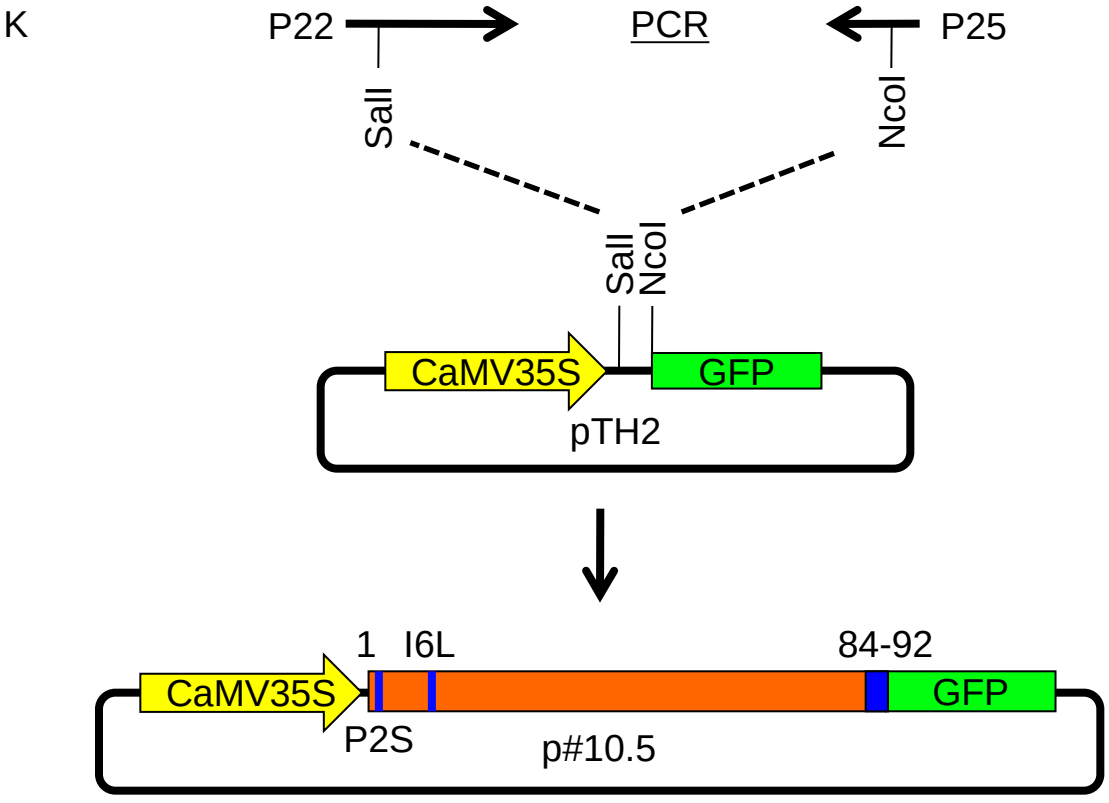

1
89
CaMV35S
GFP
p#8
G
P22
PCR
P21
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
I6L
89
CaMV35S
GFP
P2S
p#7

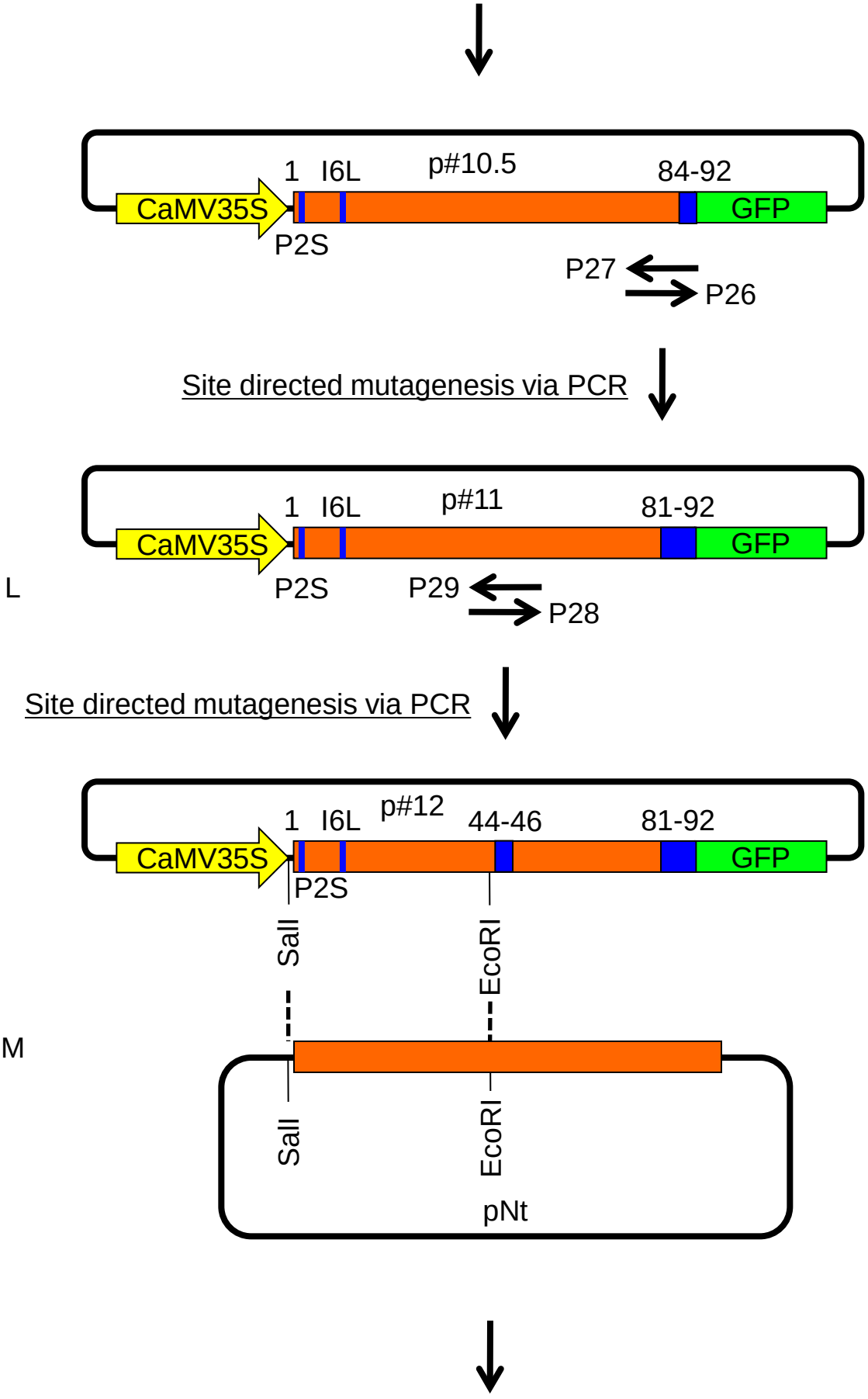

H
p#6
1
89
CaMV35S
GFP
P23
PCR
P21
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
89
CaMV35S
GFP
P2S
p#8
I
p#6
1
89
CaMV35S
GFP
P24
PCR
P21
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
89
CaMV35S
GFP
I6L
p#9

J


p#2
1
94
CaMV35S
GFP
P22
PCR
P17
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
I6L
94
CaMV35S
GFP
P2S
p#10

K


P22
PCR
P25
SalI
NcoI
SalI
NcoI
CaMV35S
GFP
pTH2
1
I6L
84-92
CaMV35S
GFP
P2S
p#10.5


1
I6L
p#10.5
84-92
CaMV35S
GFP
P2S
P27
P26
Site directed mutagenesis via PCR
1
I6L
p#11
81-92
CaMV35S
GFP
L
P2S
P29
P28
Site directed mutagenesis via PCR
1
I6L
p#12
44-46
81-92
CaMV35S
GFP
P2S
SalI
EcoRI
M
SalI
EcoRI
pNt

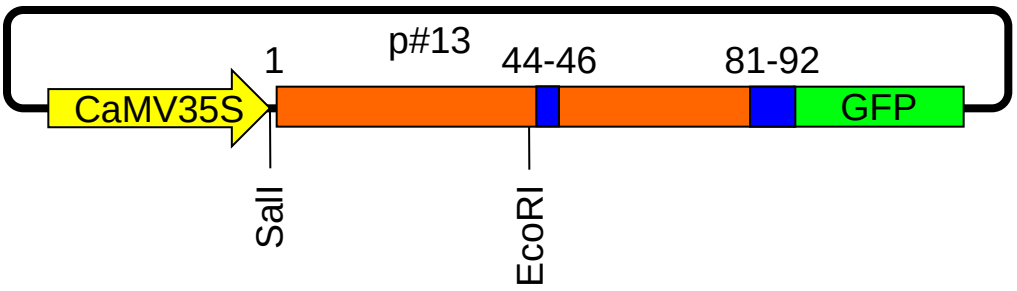


Fig. S1 Schematic illustration of plasmid construction. CaMV35S indicates 35S promoter region of Cauliflower Mosaic Virus. Primers are shown by horizontal arrows. Colors of the gene-coding regions are the same as in Fig. 4. Restriction sites are shown when necessary. See Table S1 for details. A to M corresponds to the procedures for construction of #1 to #13, respectively.

```
   1  GAAAGACTTCTCAATTCTCTCTCCTCTCTCCTGAATTAGGGTTTAGAACTCGTCAACTCG  60
  61  GCTGAGTGCACAGTCGCACACCGCACCACTCGCCTCTCTCCTCTCGCAATCCGAGTCTCC 120
 121  GACGCCGGTAAGCATTGCTCGCATTCCGACAGTTCGACAGTTCACAGTCGCAAACTCGCA 180
 181  ACTCGCATTCTCCGATTCTCGCACGAAGCATTGCTCATTTCTCCTCTTCTGTCTTCTACT 240
 241  CGAATTTGCTTAGGTAATTGACTAATTTATGATTTAATTGCTTAAACTTTCATACCAATT 300
 301  CATTTAGTAATTAGTAATCAAGTAATGCTTGGATAGATGATTATTTACAAATTAAATTAA 360
 361  TGGAATGCTTGAATGAATTGAAATTATTGTCTTATTGATTGCCTCATTGCCATGAGATTA 420
 421  ATTGAGACTCTCAAGTTTCAAATTTGTTGGGTGAATTTGTTATTGTTATTGAAGTTTAAT 480
 481  TTGGATTTTGGGTGTTGTTATTGGTTATTGAAGAGAACTTTGTTGAATTGAATTATTCAT 540
 541  CATTAGAAATTAGAGATTTAGAGTATGATTTAGTGATGTCAAATTCAATTTAGTGATGTT 600
 601  TTACTCAATGACTTATGTGAAGAACAAGTTGCGTAATAGCATGGGTGATCAACTATTGAG 660
 661  TAATGGTTTAGTCACCTTTCTTGAGAGAGATCTTTTTGCTCGAGTTAGTGATGATGATGT 720
 721  ATTAGAATGCTTTCAAAATATGAAAACTCGTAGAATGATTTTTTTAGTTTTTTTTCTTGT 780
 781  AATTTAAGAAATAATTCTCGTATTTATCGGTTGTAATTCCGAACTTGAATTGATGTATTT 840
 841  TATTTTTAAGCAATTTATATTGGTCGTCTCGTTAATTCCGAACCCACGTGGGTGAAATCC 900
 901  TGGATACGCCACTGATGAGGAGTAATGTATTTCTAACTGTTATTGTAATGAAGAAAATTT 960
 961  GGGAAGAAACTTGGGGGAAATTACGGGGTTTTCGGAAGAGCGTAAGAAATGTAAGTGATT 1020
1021  TAGTTTGTTGTAAGTTATAACAGTTTGAAAATTATGAGTTAGGTTTGGTGAAGTTGCTTG 1080
1081  ATTAAGCAGTTTATAGTGTTATTTGGTTGAGGGTGTTTGTGTTTTTTTATGTGTGCTATC 1140
1141  ATTCCGGATGTTGATTATAAACGCCTCCGCTGACACCGTGAGGAGGTTTGAGTACTAGCA 1200
1201  ATGTTGTTCTTCTTTAGATTGTTTGAGGAAAGCTTTGATATAAGATCGAATTTGGGGAGT 1260
1261  AAGTTTGAGATAGTAGAAGTGGGAGCTTATAGTTATCACTCCCTTGTTTTGCTTGTAGGC 1320
1321  TTTAGATAATTAGAATTGGAATTAGCTTTATCTCCTTTGTAGTTATTGTACCGAAACTCG 1380
1381  CATTAGTCATCTTATACTAAGATAGGAGCTTTCACTTCTTTCCCTCTCCACCTCATCCTA 1440
1441  AAATTTCATAATAGCAACCACACCAATTTGATTGATTCTTTTTAAAAGTATGCAATATTT 1500
1501  GATTCTTGTGACAGAGAAATTGATTCTCTACAAAATTCAAAGAAGCTAGTAAAATAACTT 1560
1561  ACACTCACTGAAGTACTAAGTAGGAACGAGGTTTTTGGGGAAAAAGGAGCAAAGGAGGGG 1620
1621  AGCTATGAAAGTGGGGATCAGGAAATCGAACCCTATTTTCTTTGTGTTGCAGTTTTCTAC 1680
1681  TCATTCAGCCGTTTTCGCACTCCATCCCCAAGTCCCCGACCCTTGTCTCTTTACATTTCA 1740
1741  CCATTTCCAGATTCATCCCCGAGTGTCAATTGAAGGTTTAAGTTTTTGAAGTGACAGCCC 1800
1801  TCGTCCACCTCATGAACTCAATCGAAGATGCAATCACTCTGCTCTTTAACAAGGCTCAAC 1860
1861  TCACTCAAATATCTCGAATATTGAAGCCTAAATCGAAAATATGTATGGGAATTGGGGACC 1920
1921  TCTATAATTGAAGACTTGGTTTGTTTATCTTTTGGTTCTTGTTTGAATAAGCGTATGAGG 1980
1981  TTTCCAATAATATGAGACACGCTATTTTGAATTTTCCTTGTTTTCATTCCCTTCTCAGCA 2040
2041  TTGAACCTGCTGATTATTGCCTACTACATTGCACAAAAGAGTCGAGAAATATGCCTGTAT 2100
2101  TCGTTGTTGTGCGTATCAGATTATGGAAATTCATGGCATACGACTGCGGTCTTCATTTGT 2160
2161  CAAACTTAGCTTGTTGGAGCATTGTATAGCCAAAGTCTTGCTATCGCTATTGGCCTATTT 2220
2221  GTATTGGGGGGTATGAAGTTAACAATAATCTTGAATTGCTCACATTTAATATGTCCACAT 2280
2281  CTAGTAATACACTAAACACCATCTGCCCTTGACTCTTCTTCGAAACCTATGGTGGAAGAT 2340
2341  TTTGTTGTGAATATTGGTGAAGATATCATGAAATGGTGGATGATGAAAGTTGACGTTGAA 2400
2401  CAAGTGTGGGACTTTCAGTATCTTTTTTTTTATGTTTTCAGATGGGAGAAGATTGAAACT 2460
2461  GTCCTGATATTTTTATTAAATTCAGGGTTGTTACTTTCTGTTGATTTTGTTGTTCAGCAT 2520
2521  AATTTATTATAAAATGGAAATTAAGAAATTAAGTAAAATTAGTTATAATTAGTAGCGGGA 2580
2581  GAAAAAACAAGTTTTTGGGTTTTGCCTGCTAGGCTTAAGTTCATGCTCTGGATGAGTTAC 2640
2641  ATGGTGAGCTTTTGCGTTCAAGTTTCGGAATTGGCGTATGATTGTCAAGTTTGTCAATTT 2700
2701  CCTCGATGTAAAGATACTTGGCCTTGGATCAATGCAATGGTTGATTTAATGCTTACTTTA 2760
2761  CTGTTTTCCATGGCTATTTTACATTGAGTTTGTACATCTATTCTCCACGTCCATTCACAA 2820

2821  TTTATGCATTGTTACAGAGTGACAATTCCGGAATCATGTCTCGTCGATCATTATGGAAAG 2880
                                      M  S  R  R  S  L  W  K  G

2881  GTTCTTTTGTTGATGCATTTTTGCTGAGACTGAAGGAGAACAAAGAGCAGCTTGCTAACA 2940
        S  F  V  D  A  F  L  L  R  L  K  E  N  K  E  Q  L  A  N  K

2941  AAAAGATCTGGTCAAGGAGGTCTTCTATCTTGCCAGAATTTGTTGGTAGCGCTGTTCGAA 3000
        K  I  W  S  R  R  S  S  I  L  P  E  F  V  G  S  A  V  R  I
```

```
3001  TTTACAATGGAAAGTCATTCGTTCGCTGTAAGATCACTGAATCAAAAGTTGGTCACAAAT 3060
        Y  N  G  K  S  F  V  R  C  K  I  T  E  S  K  V  G  H  K  F

3061  TTGGAGAGTTTGCTTTCACACGGAAAAGAAGGCCTCACAAATTAGATGGTGGAAAGAAAA 3120
        G  E  F  A  F  T  R  K  R  R  P  H  K  L  D  G  G  K  K  K

3121  AGGGGGGTCGTTGAGCTTGCCTCAGCTGCCTTCATTTACTGTGCTTCCGTGTTCAATTTT 3180
        G  G  R  *

3181  CTGGAGGGTACTTTGCTTCTTGCTTTCTTCTTGTTGAACTAGTTTTTATGCCATGACAAA 3240
3241  TAACACTGTGTTTAAGGGAGTTTTTATGAAATCCCTAAAAAACCTGTATCAAATTTGATT 3300
3301  CGATGACTTATGCCTTACTAATTATTGCTTAGCTAAAAGGGCCATTTTTTGTGATTCGAT 3360
3361  TTTGATCATCATATTTACAACTTTTATCTCCTATATTCAGTGGAAGTATTGGGATTTGGG 3420
3421  AACATCATATATAGTTGATGCTTGTACATTGCAGGATTATATCGATTAAATTTTCATGTG 3480
3481  ATATTCCTTTCCCAAATTTTAAAATTGGCTCAGGGTTCACACCTAGTAGAATCGAAATGC 3540
3541  CTCTTTGTATGTTTCGTGGAGAACAGTTCTCTAAATAGCTGGAATCTTATGGTAATTTTC 3600
3601  TCAGTTGAATATGTTTCAACGATCATTGCAGAGGATTAGATCTTTGATAGAATTGGATGA 3660
3661  CCTAAGTTTGGTGGGTACATCAAACATGAATCAAATCTAGGTATCTAAAATGGTTGTTGA 3720
3721  ACAGTCCAACTCAAAGGGGTCTTCGAATGACTTGACAAAGAGCTTCTAAATACATTTTTC 3780
3781  TAATTCAGATGCTGACCTATCAGTTGAGAATGTTTGACGAAGCTT                3825
```

Fig. S2 Nucleotide sequence of sugar beet *Rps19* locus. The open reading frame was translated into amino acid residues (shown in one letter symbols below the nucleotide sequence). Untranslated regions are shown by dashed underlines. Donor- (5'-GT-3') and acceptor sites (5'-AG-3') of the intron are boxed. An in-frame stop codon is indicated by red-colored letters.

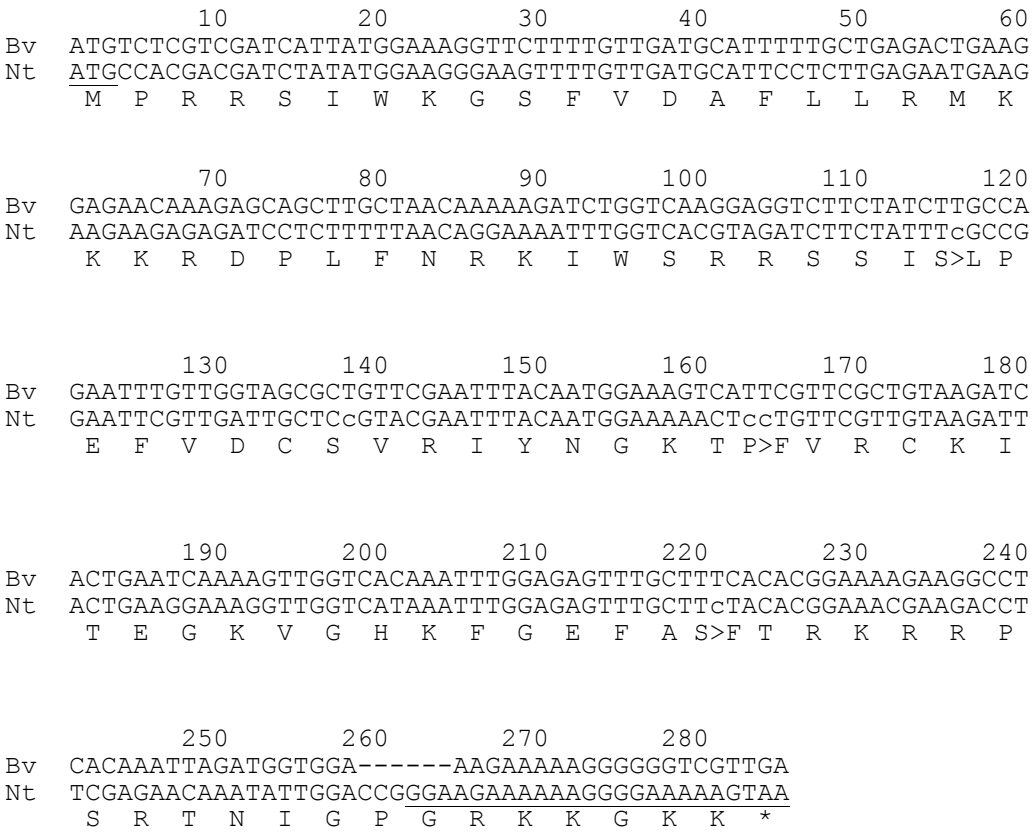
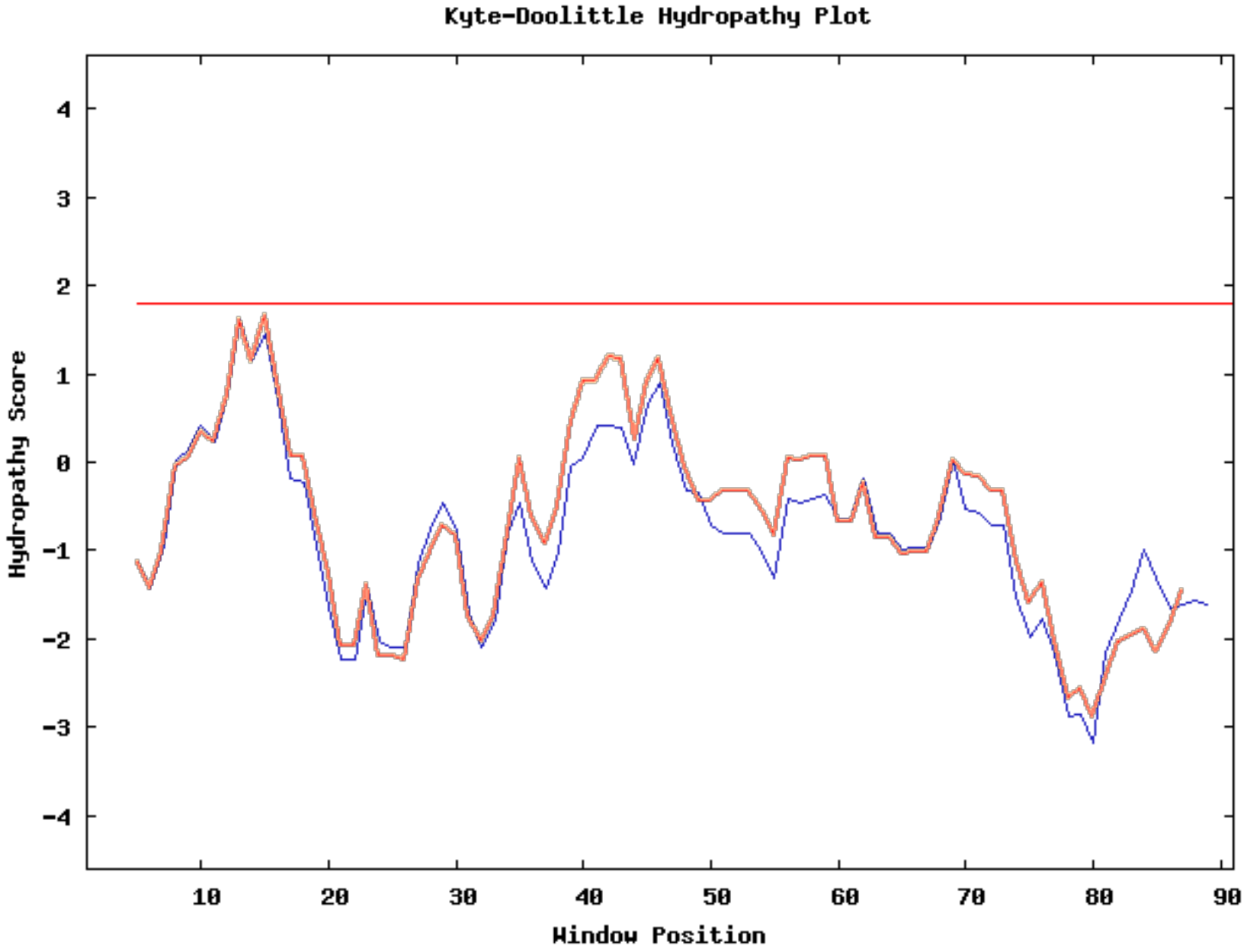


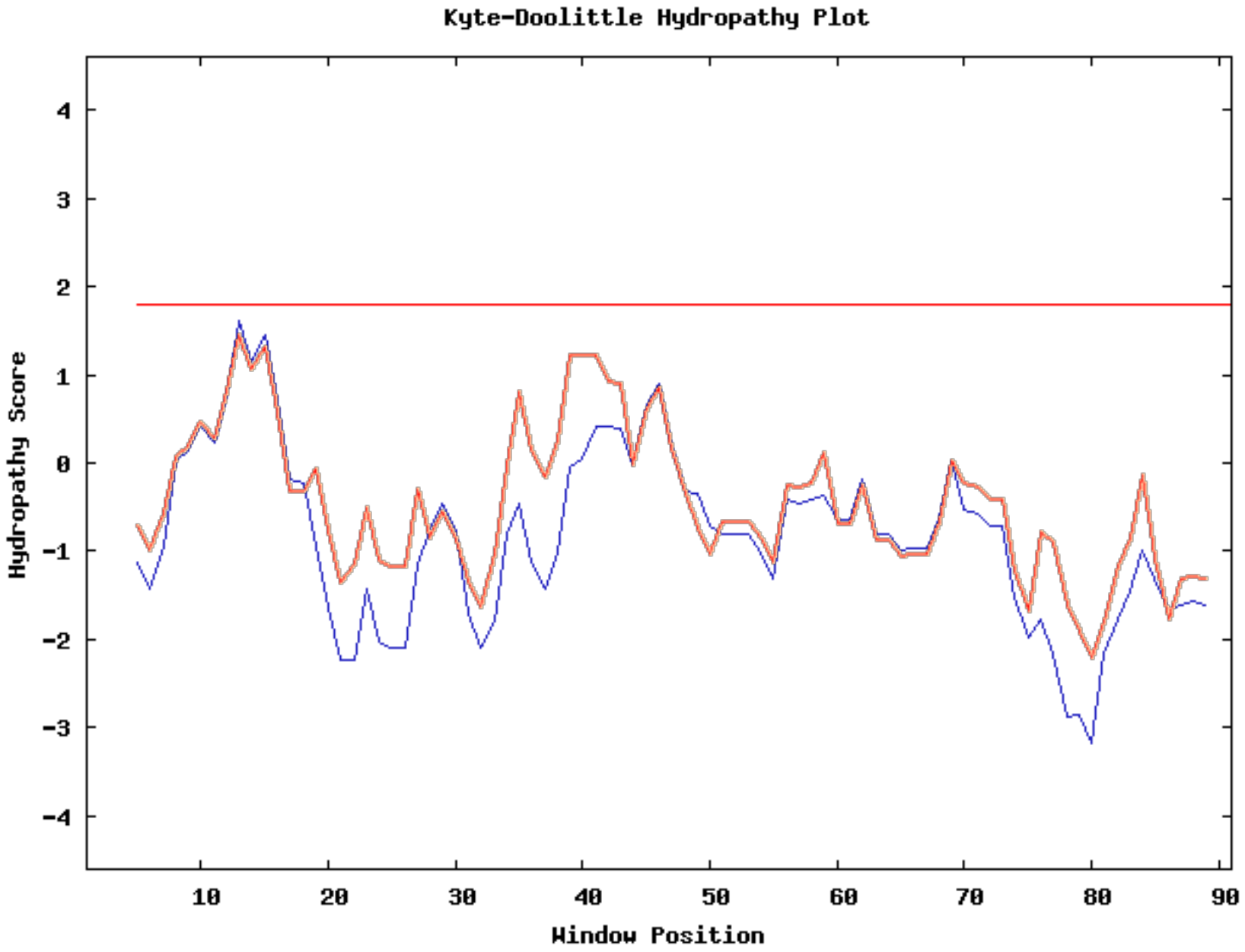
Fig. S3 Alignment of nucleotide sequences of sugar-beet *Rps19* (Bv) and tobacco *rps19* (Nt). Amino acid sequence of putative translation product of the tobacco *rps19* is shown below. Primer sequences are shown by underlined font. RNA editing sites are shown by lower-case letters. Alteration of amino acid specificity by RNA editing is shown below the editing sites.

A



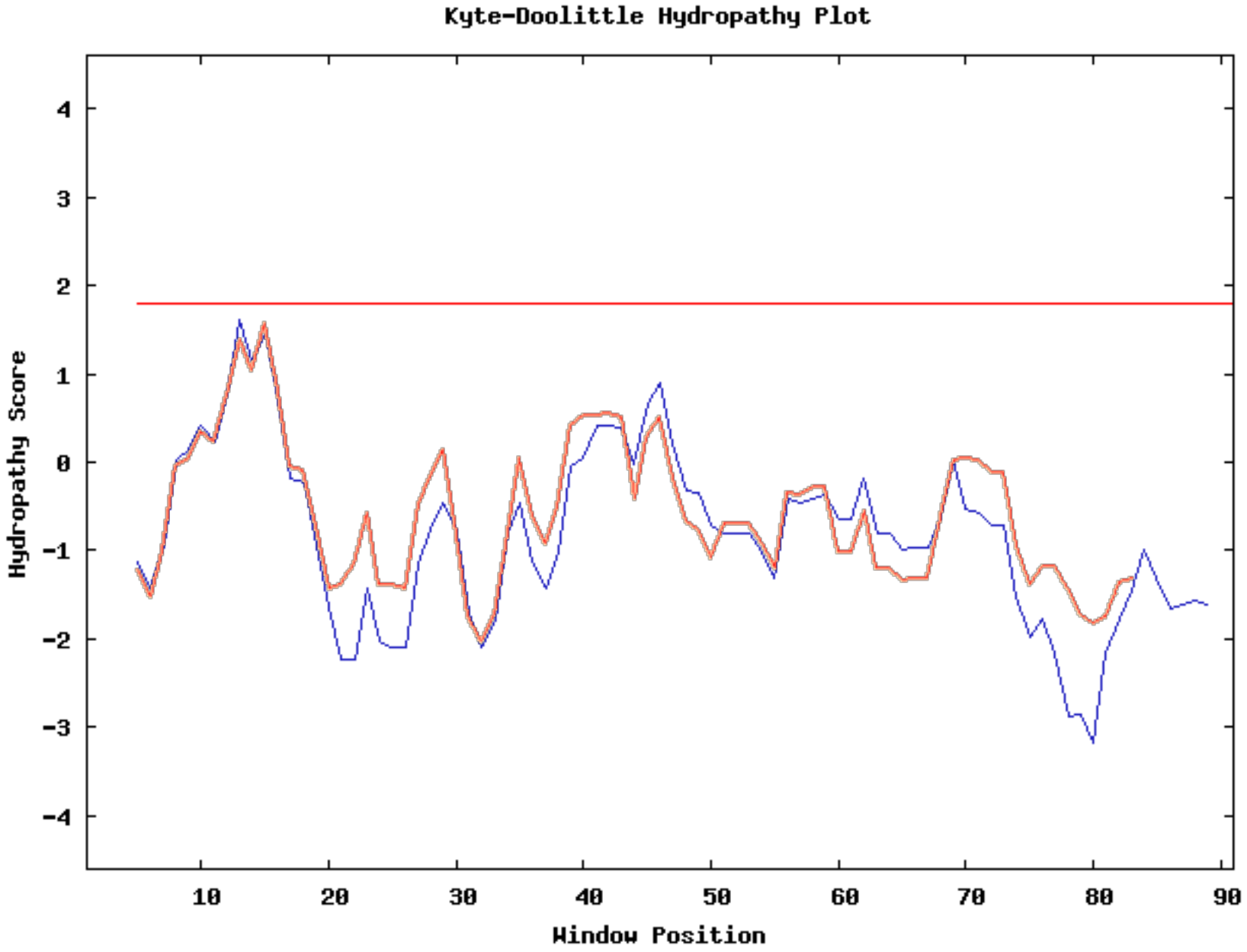
NtRPS19 (blue) vs. BvRPS19 (red)

B



NtRPS19 (blue) vs. melon RPS19 (red)

C



NtRPS19 (blue) vs. orange RPS19 (red)

Fig. S4 Comparison of Kyte-Doolittle hydropathy plots of NtRPS19 with three ITS-type RPS19 proteins, BvRPS19, melon RPS19 and orange RPS19. Plots were drawn at the web site (http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm) with default parameters. The two images were merged using PowerPoint (Microsoft). Panels A, B, and C shows a comparison between NtRPS19 and BvRPS19, NtRPS19 and melon RPS19, and NtRPS19 and orange RPS19, respectively.

Supporting information-1 Screening the genomic library and identifying the *rps19*-like ORF

To see the molecular organization of the region that had homology to petunia *rps19*, we screened a sugar-beet genomic library consisting of recombinant lambda phages. We obtained three positive clones that overlapped with each other and constituted a 14.5-kbp contig according to restriction mapping.

A 2.6-kbp *Hin*dIII fragment, which appeared to correspond to the 2.6-kbp *Hin*dIII band on the blot, was subcloned from one of the three recombinant phage DNAs into a plasmid vector, and was sequenced entirely. We obtained a 2596-bp sequence that contained an ORF encoding 92 amino acid-residue polypeptide showing high homology (62-75% identity) to land plant mitochondrial *rps19* (Fig. S2 and Fig. 3). No other homologous sequence was retrieved from DDBJ/EMBL/GenBank database by BLASTN and BLASTX.